NOTES ON PERELMAN'S PAPERS

BRUCE KLEINER AND JOHN LOTT

1. Introduction

These are notes on Perelman's papers "The Entropy Formula for the Ricci Flow and its Geometric Applications" [51] and "Ricci Flow with Surgery on Three-Manifolds' [52]. In these two remarkable preprints, which were posted on the ArXiv in 2002 and 2003, Grisha Perelman announced a proof of the Poincaré Conjecture, and more generally Thurston's Geometrization Conjecture, using the Ricci flow approach of Hamilton. Perelman's proofs are concise and, at times, sketchy. The purpose of these notes is to provide the details that are missing in [51] and [52], which contain Perelman's arguments for the Geometrization Conjecture.

Among other things, we cover the construction of the Ricci flow with surgery of [52]. We also discuss the long-time behavior of the Ricci flow with surgery, which is needed for the full Geometrization Conjecture. The papers [24, 53], which are not covered in these notes, each provide a shortcut in the case of the Poincaré Conjecture. Namely, these papers show that if the initial manifold is simply-connected then the Ricci flow with surgery becomes extinct in a finite time, thereby removing the issue of the long-time behavior. Combining this claim with the proof of existence of Ricci flow with surgery gives the shortened proof in the simply-connected case.

These notes are intended for readers with a solid background in geometric analysis. Good sources for background material on Ricci flow are [22, 23, 33, 66]. The notes are self-contained but are designed to be read along with [51, 52]. For the most part we follow the format of [51, 52] and use the section numbers of [51, 52] to label our sections. We have done this in order to respect the structure of [51, 52] and to facilitate the use of the present notes as a companion to [51, 52]. In some places we have rearranged Perelman's arguments or provided alternative arguments, but we have refrained from an overall reorganization.

Besides providing details for Perelman's proofs, we have included some expository material in the form of overviews and appendices. Section 3 contains an overview of the Ricci flow approach to geometrization of 3-manifolds. Sections 4 and 57 contain overviews of [51] and [52], respectively. The appendices discuss some background material and techniques that are used throughout the notes.

Regarding the proofs, the papers [51, 52] contain some incorrect statements and incomplete arguments, which we have attempted to point out to the reader. (Some of the mistakes in [51] were corrected in [52].) We did not find any serious problems, meaning problems that cannot be corrected using the methods introduced by Perelman.

Date: September 19, 2008.
Research supported by NSF grants DMS-0306242 and DMS-0204506, and the Clay Mathematics Institute.

We will refer to Section X.Y of [51] as I.X.Y, and Section X.Y of [52] as II.X.Y. A reader may wish to start with the overviews, which explain the logical structure of the arguments and the interrelations between the sections. It may also be helpful to browse through the appendices before delving into the main body of the material.

These notes have gone through various versions, which were posted at [39]. An initial version with notes on [51] was posted in June 2003. A version covering [51, 52] was posted in September 2004. After the May 2006 version of these notes was posted on the ArXiv, expositions of Perelman's work appeared in [15] and [45].

Acknowledgements. In the preparation of the September 2004 version of our notes, we benefited from a workshop on Perelman's surgery procedure that was held in August 2004, at Princeton. We thank the participants of that meeting, as well as the Clay Mathematics Institute for supporting it.

We are grateful for comments and corrections that we have received regarding earlier versions of these notes. We especially thank Gérard Besson for comments on Theorem 13.3 and Lemma 80.4, Peng Lu for comments on Lemma 74.1 and Bernhard Leeb for discussions on the issue of uniqueness for the standard solution. Along with these people we thank Mike Anderson, Albert Chau, Ben Chow, Xianzhe Dai, Jonathan Dinkelbach, Sylvain Maillot, John Morgan, Joan Porti, Gang Tian, Peter Topping, Bing Wang, Guofang Wei, Hartmut Weiss, Jon Wolfson, Rugang Ye and Maciej Zworski.

We thank the referees for their detailed and helpful comments on this paper.

We thank the Clay Mathematics Institute for supporting the writing of the notes on [52].

Contents

2. A reading guide

Perelman's papers contain a wealth of results about Ricci flow. We cover all of these results, whether or not they are directly relevant to the Poincaré and Geometrization Conjectures.

Some readers may wish to take an abbreviated route that focuses on the proof of the Poincaré Conjecture or the Geometrization Conjecture. Such readers can try the following itinerary.

Begin with the overviews in Sections 3 and 4. Then review Hamilton's compactness theorem and its variants, as described in Appendix E; an exposition is in [66, Chapter 7]. Next, read I.7 (Sections 15-26), followed by I.8.3(b) (Section 27). After reviewing the theory of Riemannian manifolds and Alexandrov spaces of nonnegative sectional curvature (Appendix G and references therein), proceed to I.11 (Sections 38-50), followed by II.1.2 and I.12.1 (Sections 51-52).

At this point, the reader should be ready for the overview of Perelman's second paper in Section 57, and can proceed with II.1-II.5 (Sections 58-80). In conjunction with one of the finite extinction time results [24, 25, 53], this completes the proof of the Poincaré Conjecture.

To proceed with the rest of the proof of the Geometrization Conjecture, the reader can begin with the large-time estimates for nonsingular Ricci flows, which appear in I.12.2-I.12.4 (Sections 53-55). The reader can then go to II.6 and II.7 (Sections 81-92).

The main topics that are missed by such an abbreviated route are the $\mathcal{F}$ and $\mathcal{W}$ functionals (Sections 5-14), Perelman's differential Harnack inequality (Section 29), pseudolocality (Sections 30-37) and Perelman's alternative proof of cusp incompressibility (Section 93).

3. An overview of the Ricci flow approach to 3-manifold geometrization

This section is an overview of the Ricci flow approach to 3-manifold geometrization. We make no attempt to present the history of the ideas that go into the argument. We caution the reader that for the sake of readability, in many places we have suppressed technical points and deliberately oversimplified the story. The overview will introduce the argument in three passes, with successively greater precision and detail : we start with a very crude sketch, then expand this to a step-by-step outline of the strategy, and then move on to more detailed commentary on specific points.

Other overviews may be found in [14, 44]. The primary objective of our exposition is to prepare the reader for a more detailed study of Perelman's work.

We refer the reader to Appendix I for the statement of the geometrization conjecture.

By convention, all manifolds and Riemannian metrics in this section will be smooth. We follow the notation of [51, 52] for pointwise quantities: R denotes the scalar curvature, Ric the Ricci curvature, and $|\operatorname{Rm}|$ the largest absolute value of the sectional curvatures. An inequality such as $\operatorname{Rm} \geq C$ means that all of the sectional curvatures at a point or in a region, depending on the context, are bounded below by C. In this section, we will specialize to three dimensions.

3.1. **The definition of Ricci flow, and some basic properties.**

Let M be a compact 3-manifold and let $\{g(t)\}_{t\in[a,b]}$ be a smoothly varying family of Riemannian metrics on M. Then $g(\cdot)$ satisfies the *Ricci flow equation* if

$$\frac{\partial g}{\partial t}(t) \;=\; -\,2\ \operatorname{Ric}(g(t)) \tag{3.1}$$

holds for every $t \in [a, b]$. Hamilton showed in [35] that for any Riemannian metric g_0 on M, there is a $T \in (0, \infty]$ with the property that there is a (unique) solution $g(\cdot)$ to the Ricci flow equation defined on the time interval $[0, T)$ with $g(0) = g_0$, so that if $T < \infty$ then the curvature of $g(t)$ becomes unbounded as $t \to T$. We refer to this maximal solution as the Ricci flow with initial condition g_0. If $T < \infty$ then we call T the blow-up time. A basic example is the shrinking round 3-sphere, with $g_0 = r_0^2\, g_{S^3}$ and $g(t) = (r_0^2 - 4t)\, g_{S^3}$, in which case $T = \frac{r_0^2}{4}$.

Suppose that M is simply-connected. Based on the round 3-sphere example, one could hope that every Ricci flow on M blows up in finite time and becomes round while shrinking to a point, as t approaches the blow-up time T. If so, then by rescaling and taking a limit as $t \to T$, one would show that M admits a metric of constant positive sectional curvature and therefore, by a classical theorem, is diffeomorphic to S^3. The analogous argument does work in two dimensions [22, Chapter 5]. Furthermore, if the initial metric g_0 has positive Ricci curvature then Hamilton showed in [33] that this is the correct picture: the manifold shrinks to a point in finite time and becomes round as it shrinks.

One is then led to ask what can happen if M is simply-connected but g_0 does not have positive Ricci curvature. Here a new phenomenon can occur — the Ricci flow solution may become singular before it has time to shrink to a point, due to a possible neckpinch. A neckpinch is modeled by a product region $(-c, c) \times S^2$ in which one or many S^2-fibers

separately shrink to a point at time T, due to the positive curvature of S^2. The formation of neckpinch (and other) singularities prevents one from continuing the Ricci flow. In order to continue the evolution some intervention is required, and this is the role of surgery. Roughly speaking, the idea of surgery is to remove a neighborhood diffeomorphic to $(-c', c') \times S^2$ containing the shrinking 2-spheres, and cap off the resulting boundary components by gluing in 3-balls. Of course the topology of the manifold changes during surgery – for instance it may become disconnected – but it changes in a controlled way. The postsurgery Riemannian manifold is smooth, so one may restart the Ricci flow using it as an initial condition. When continuing the flow one may encounter further neckpinches, which give rise to further surgeries, etc. One hopes that eventually all of the connected components shrink to points while becoming round, i.e. that the Ricci flow solution has a *finite extinction time*.

3.2. **A rough outline of the Ricci flow proof of the Poincaré Conjecture.** We now give a step-by-step glimpse of the proof, stating the needed steps as claims.

One starts with a compact orientable 3-manifold M with an arbitrary metric g_0. For the moment we do not assume that M is simply-connected. Let $g(\cdot)$ be the Ricci flow with initial condition g_0, defined on $[0, T)$. Suppose that $T < \infty$. Let $\Omega \subset M$ be the set of points $x \in M$ for which $\lim_{t \to T^-} R(x, t)$ exists and is finite. Then $M - \Omega$ is the part of M that is going singular. (For example, in the case of a single standard neckpinch, $M - \Omega$ is a 2-sphere.) The first claim says what M looks like near this singularity set.

Claim 3.2. [52] *The set Ω is open and as $t \to T$, the evolving metric $g(\cdot)$ converges smoothly on compact subsets of Ω to a Riemannian metric $\overline{g}$. There is a geometrically defined neighborhood U of $M - \Omega$ such that each connected component of U is either*

A. Compact and diffeomorphic to $S^1 \times S^2$, $S^1 \times_{\mathbb{Z}_2} S^2$ or S^3/Γ, where Γ is a finite subgroup of $\mathrm{SO}(4)$ that acts freely and isometrically on the round S^3. (In writing $S^1 \times_{\mathbb{Z}_2} S^2$, the generator of $\mathbb{Z}_2$ acts on S^1 by complex conjugation and on S^2 by the antipodal map. Then $S^1 \times_{\mathbb{Z}_2} S^2$ is diffeomorphic to $\mathbb{R}P^3 \# \mathbb{R}P^3$.)

or

B. Noncompact and diffeomorphic to $\mathbb{R} \times S^2$, $\mathbb{R}^3$ or the twisted line bundle $\mathbb{R} \times_{\mathbb{Z}_2} S^2$ over $\mathbb{R}P^2$.

In Case B, the connected component meets Ω in geometrically controlled collar regions diffeomorphic to $\mathbb{R} \times S^2$.

Thus Claim 3.2 provides a topological description of a neighborhood U of the region $M - \Omega$ where the Ricci flow is going singular, along with some geometric control on U.

Claim 3.3. [52] *There is a well-defined way to perform surgery on M, which yields a smooth post-surgery manifold M' with a Riemannian metric g'.*

Claim 3.3 means that there is a well-defined procedure for specifying the part of M that will be removed, and for gluing caps on the resulting manifold with boundary. The discarded part corresponds to the neighborhood U in Claim 3.2. The procedure is required to satisfy a number of additional conditions which we do not mention here.

Undoing the surgery, i.e. going from the postsurgery manifold to the presurgery manifold, amounts to restoring some discarded components (as in Case A of Claim 3.2) and performing connected sums of some of the components of the postsurgery manifold, along with some possible connected sums with a finite number of new $S^1 \times S^2$ and $\mathbb{R}P^3$ factors. The $S^1 \times S^2$ comes from the case when a surgery does not disconnect the connected component where it is performed. The $\mathbb{R}P^3$ factors arise from the twisted line bundle components in Case B of Claim 3.2.

After performing a surgery one lets the new manifold evolve under the Ricci flow until one encounters the next blowup time (if there is one). One then performs further surgery, lets the new manifold evolve, and repeats the process.

Claim 3.4. [52] *One can arrange the surgery procedure so that the surgery times do not accumulate.*

If the surgery times were to accumulate, then one would have trouble continuing the flow further, effectively killing the whole program. Claim 3.4 implies that by alternating Ricci flow and surgery, one obtains an evolutionary process that is defined for all time (though the manifold may become the empty set from some time onward). We call this *Ricci flow with surgery.*

Claim 3.5. [24, 25, 53] *If the original manifold M is simply-connected then any Ricci flow with surgery on M becomes extinct in finite time.*

Having a finite extinction time means that from some time onwards, the manifold is the empty set. More generally, the same proof shows that if the prime decomposition of the original manifold M has no aspherical factors, then every Ricci flow with surgery on M becomes extinct in finite time. (Recall that a connected manifold X is aspherical if $\pi_k(X) = 0$ for all $k > 1$ or, equivalently, if its universal cover is contractible.)

The Poincaré Conjecture follows immediately from the above claims. From Claim 3.5, after some finite time the manifold is the empty set. From Claims 3.2, 3.3 and 3.4, the original manifold M is diffeomorphic to a connected sum of factors that are each $S^1 \times S^2$ or a standard quotient S^3/Γ of S^3. As we are assuming that M is simply-connected, van Kampen's theorem implies that M is diffeomorphic to a connected sum of S^3's, and hence is diffeomorphic to S^3.

3.3. **Outline of the proof of the Geometrization Conjecture.** We now drop the assumption that M is simply-connected. The main difference is that Claim 3.5 no longer applies, so the Ricci flow with surgery may go on forever in a nontrivial way. (We remark that Claim 3.5 is needed only for a shortened proof of the Poincaré Conjecture; the proof in the general case is logically independent of Claim 3.5 and also implies the Poincaré Conjecture.) The possibility that there are infinitely many surgery times is not excluded, although it is not known whether this can actually happen.

A simple example of a Ricci flow that does not become extinct is when $M = H^3/\Gamma$, where Γ is a freely-acting cocompact discrete subgroup of the orientation-preserving isometries of hyperbolic space H^3. If g_{hyp} denotes the metric on M of constant sectional curvature -1

and $g_0 = r_0^2 g_{hyp}$ then $g(t) = (r_0^2 + 4t)\, g_{hyp}$. Putting $\widehat{g}(t) = \frac{1}{t} g(t)$, one finds that $\lim_{t\to\infty} \widehat{g}(t)$ is the metric on M of constant sectional curvature $-\frac{1}{4}$, independent of r_0.

Returning to the general case, let M_t denote the time-t manifold in a Ricci flow with surgery. (If t is a surgery time then we consider the postsurgery manifold.) If for some t a component of M_t admits a metric with nonnegative scalar curvature then one can show that the component becomes extinct or admits a flat metric; either possiblity is good enough when we are trying to prove the Geometrization Conjecture for the initial manifold M. So we will assume that for every t, each component of M_t has a point with strictly negative scalar curvature.

Motivated by the hyperbolic example, we consider the metric $\widehat{g}(t) = \frac{1}{t} g(t)$ on M_t. Given $x \in M_t$, define the *intrinsic scale* $\rho(x,t)$ to be the radius ρ such that $\inf_{B(x,\rho)} \mathrm{Rm} = -\rho^{-2}$, where Rm denotes the sectional curvature of $\widehat{g}(t)$; this is well-defined because the scalar curvature is negative somewhere in the connected component of M_t containing x. Given $w > 0$, define the *w-thick part* of M_t by

$$M^+(w,t) = \{x \in M_t : \mathrm{vol}(B(x,\rho(x,t))) > w\, \rho(x,t)^3\}. \tag{3.6}$$

It is not excluded that $M^+(w,t) = M_t$ or $M^+(w,t) = \emptyset$. The next claim says that for any $w > 0$, as time goes on, $M^+(w,t)$ approaches the w-thick part of a manifold of constant sectional curvature $-\frac{1}{4}$.

Claim 3.7. [52] *There is a finite collection $\{(H_i, x_i)\}_{i=1}^k$ of complete pointed finite-volume 3-manifolds with constant sectional curvature $-\frac{1}{4}$ and, for large t, a decreasing function $\alpha(t)$ tending to zero and a family of maps*

$$f_t : \coprod_{i=1}^k H_i \supset \coprod_{i=1}^k B\left(x_i, \frac{1}{\alpha(t)}\right) \to M_t\,, \tag{3.8}$$

such that
1. f_t is $\alpha(t)$-close to being an isometry.
2. The image of f_t contains $M^+(\alpha(t),t)$.
3. The image under f_t of a cuspidal torus of $\{H_i\}_{i=1}^k$ is incompressible in M_t.

The proof of Claim 3.7 uses earlier work by Hamilton [34].

Claim 3.9. [52, 64] *Let Y_t be the truncation of $\bigcup_{i=1}^k H_i$ obtained by removing horoballs at distance approximately $\frac{1}{2\alpha(t)}$ from the basepoints x_i. Then for large t, $M_t - f_t(Y_t)$ is a graph manifold.*

Claim 3.9 reduces to a statement in Riemannian geometry about 3-manifolds that are locally volume-collapsed with a lower bound on sectional curvature.

Claims 3.7 and 3.9, along with Claims 3.2-3.4, imply the geometrization conjecture, cf. Appendix I.

In the remainder of this section, we will discuss some of the claims in more detail.

3.4. **Claim 3.2 and the structure of singularities.** Claim 3.2 is derived from a more localized statement, which says that near points of large scalar curvature, the Ricci flow looks very special : it is well-approximated by a special kind of model Ricci flow, called a κ-solution.

Claim 3.10. *Suppose that we have a given Ricci flow solution on a finite time interval. If $x \in M$ and the scalar curvature $R(x,t)$ is large then in the time-t slice, there is a ball centered at x of radius comparable to $R(x,t)^{-\frac{1}{2}}$ in which the geometry of the Ricci flow is close to that of a ball in a κ-solution.*

The quantity $R(x,t)^{-\frac{1}{2}}$ is sometimes called the *curvature scale* at (x,t), because it scales like a distance. We will define κ-solutions below, but mention here that they are Ricci flows with nonnegative sectional curvature, and they are *ancient*, i.e. defined on a time interval of the form $(-\infty, a)$.

The strength of Claim 3.10 comes from the fact that there is a good description of κ-solutions.

Claim 3.11. [52] *Any three-dimensional oriented κ-solution $(M_\infty, g_\infty(\cdot))$ falls into one of the following types :*
(a) A finite isometric quotient of the round shrinking 3-sphere.
(b) A Ricci flow on a manifold diffeomorphic to S^3 or $\mathbb{R}P^3$.
(c) A standard shrinking round neck on $\mathbb{R} \times S^2$
(d) A Ricci flow on a manifold diffeomorphic to $\mathbb{R}^3$, each time slice of which is asymptotically necklike at infinity.
(e) The $\mathbb{Z}_2$-quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^2$ of a shrinking round neck.

Together, Claims 3.10 and 3.11 say that where the scalar curvature is large, there is a region of diameter comparable to the curvature scale where one sees either a closed manifold of known topology (cases (a) and (b)), a neck region (case (c)), a neck region capped off by a 3-ball (case (d)), or a neck region capped off by a twisted line bundle over $\mathbb{R}P^2$ (case (e)). Applying this statement to every point of large scalar curvature at a time t just prior to the blow-up time T, one obtains a cover of M by regions with special geometry and topology. Any overlaps occur in neck-like regions, permitting one to splice them together to form the connected components with known topology whose existence is asserted in Claim 3.2.

Claim 3.10 is proved using a rescaling (or blow-up) argument. This is a standard technique in geometric analysis and PDE's for treating scale-invariant equations, such as the Ricci flow equation. The claim is equivalent to the statement that if $\{(x_i, t_i)\}_{i=1}^\infty$ is a sequence of spacetime points for which $\lim_{i\to\infty} R(x_i, t_i) = \infty$, then by rescaling the Ricci flow and passing to a subsequence, one obtains a new sequence of Ricci flows which converges to a κ-solution. More precisely, view (x_i, t_i) as a new spacetime basepoint and spatially expand the solution around (x_i, t_i) by $R(x_i, t_i)^{\frac{1}{2}}$. For dimensional reasons, in order for rescaling to produce a new Ricci flow solution one must also expand the time factor by $R(x_i, t_i)$. The new Ricci flow solution, with time parameter s, is given by

$$\overline{g}_i(s) \;=\; R(x_i, t_i)\, g\left(R(x_i, t_i)^{-1}\, s \;+\; t_i\right). \tag{3.12}$$

The new time interval for s is $[-\,R(x_i, t_i)\, t_i, 0]$. One would then hope to take an appropriate limit $(M_\infty, \overline{g}_\infty)$ of a subsequence of these rescaled solutions $\{(M, \overline{g}_i(\cdot))\}_{i=1}^\infty$. (Technically

speaking, one uses smooth convergence of sequences of Ricci flows with basepoints; this notion of convergence allows us to focus on what happens near the spacetime points (x_i, t_i).) Any such limit solution $(M_\infty, \overline{g}_\infty)$ will be an ancient solution, since $\lim_{i\to\infty} -R(x_i,t_i)\, t_i = -\infty$. Furthermore, from a 3-dimensional result of Hamilton and Ivey (see Appendix B), any limit solution will have nonnegative sectional curvature.

Although this sounds promising, a major problem was to show that a limit solution actually exists. To prove this, one would like to invoke Hamilton's compactness theorem [32]. In the present situation, the compactness theorem says that the sequence of rescaled Ricci flows $\{(M, \overline{g}_i(\cdot))\}_{i=1}^\infty$ has a smoothly convergent subsequence provided two conditions are met:

A. For every $r > 0$ and sufficiently large i, the sectional curvature of $\overline{g_i}$ is bounded uniformly independent of i at each point (x, s) in spacetime such that x lies in the $\overline{g_i}$-ball $B_0(x_i, r)$ and $s \in [-r^2, 0]$ and

B. The injectivity radii $\mathrm{inj}(x_i, 0)$ in the time-0 slices of the $\overline{g_i}$'s have a uniform positive lower bound.

For the moment, we ignore the issue of verifying condition A, and simply assume that it holds for the sequence $\{(M, \overline{g}_i(\cdot))\}_{i=1}^\infty$. In the presence of the sectional curvature bounds in condition A, a lower bound on the injectivity radius is known to be equivalent to a lower bound on the volume of metric balls. In terms of the original Ricci flow solution, this becomes the condition that

$$r^{-3}\,\mathrm{vol}(B_t(x,r)) \geq \kappa \;>\; 0, \tag{3.13}$$

where $B_t(x, r)$ is an arbitrary metric r-ball in a time-t slice, and the curvature bound $|\,\mathrm{Rm}\,| \leq \frac{1}{r^2}$ holds in $B_t(x, r)$. The number κ could depend on the given Ricci flow solution, but the bound (3.13) should hold for all $t \in [0, T)$ and all $r < \rho$, where ρ is a relevant scale.

One of the outstanding achievements of [51] is to prove that for an arbitrary Ricci flow defined on a finite time interval, equation (3.13) does hold with appropriate values of κ and ρ. In fact, the proof works in arbitrary dimension. This result is called a "no local collapsing theorem" because it excludes the phenomenon of Cheeger-Gromov collapse, in which a sequence of Riemannian manifolds has uniformly bounded curvature, but fails to converge because the injectivity radii tend to zero.

One can then apply the no local collapsing theorem to the preceding rescaling argument, provided that one has the needed sectional curvature bounds, in order to construct the ancient solution $(M_\infty, \overline{g}_\infty)$. In the blowup limit the condition that $r < \rho$ goes away, and so we can say that $(M_\infty, \overline{g}_\infty)$ is κ-noncollapsed (i.e. satisfies (3.13)) at all scales. In addition, in the three-dimensional case one can show that $(M_\infty, \overline{g}_\infty)$ has bounded sectional curvature. To summarize, $(M_\infty, \overline{g}_\infty)$ is a *κ-solution*, meaning that it is an ancient Ricci flow solution with nonnegative curvature operator on each time slice and bounded sectional curvature on compact time intervals, which is κ-noncollapsed at all scales.

With the no local collapsing theorem in place, most of the proof of Claim 3.10 is concerned with showing that in the rescaling argument, we effectively have the needed curvature bounds

of condition A. The argument is a tour-de-force with many ingredients, including earlier work of Hamilton and the theory of Riemannian manifolds with nonnegative sectional curvature.

3.5. **The proof of Claims 3.3 and 3.4.** Claim 3.2 allows one to take the limit of the evolving metric $g(\cdot)$ as $t \to T$, on the open set Ω where the metric is not becoming singular. It also provides geometrically defined regions – the connected components of the open set U – which one removes during surgery. Each boundary component of the resulting manifold is a nearly round 2-sphere with a nearly cylindrical collar, because the collar regions in Case B of Claim 3.2 have a neck-like geometry. This enables one to glue in 3-balls with a standard metric, using a partition of unity construction.

The Ricci flow starting with the postsurgery metric may also go singular after a finite time. If so, one can appeal to Claim 3.2 again to perform surgery. However, the elapsed time between successive surgeries will depend on the scales at which surgeries are performed. Unless one performs the surgeries very carefully, the surgery times may accumulate.

The way to rule out an accumulation of surgery times is to arrange the surgery procedure so that a surgery at time t removes a definite amount of volume $v(t)$. That is, a surgery at time t should be performed at a definite scale $h(t)$. In order to guarantee that this is possible, one needs to establish a quantitative version of Claim 3.2 for a Ricci flow with surgery, which applies not just at the first surgery time T but also at a later surgery time T'. The output of this quantitative version can depend on the surgery time T' and the time-zero metric, but it should be independent of whether or when surgeries occur before time T'.

The general idea of the proof is similar to that of Claim 3.2, except that one has to carefully prescribe the surgery procedure in order to control the effect of the earlier surgeries. In particular, one of Perelman's remarkable achievements is a version of the no local collapsing theorem for Ricci flows with surgery.

We refer the reader to Section 57 for a more detailed overview of the proof of Claim 3.4, and for further discussion of Claims 3.7 and 3.9.

4. Overview of *The Entropy Formula for the Ricci Flow and its Geometric Applications* [51]

The paper [51] deals with nonsingular Ricci flows; the surgery process is considered in [52]. In particular, the final conclusion of [51] concerns Ricci flows that are singularity-free and exist for all positive time. It does not apply to compact 3-manifolds with finite fundamental group or positive scalar curvature.

The purpose of the present overview is not to give a comprehensive summary of the results of [51]. Rather we indicate its organization and the interdependence of its sections, for the convenience of the reader. Some of the remarks in the overview may only become clear once the reader has absorbed a portion of the detailed notes.

Sections I.1-I.10, along with the first part of I.11, deal with Ricci flow on n-dimensional manifolds. The second part of I.11, and Sections I.12-I.13, deal more specifically with Ricci flow on 3-dimensional manifolds. The main result is that geometrization holds if a

compact 3-manifold admits a Riemannian metric which is the initial metric of a smooth Ricci flow. This was previously shown in [34] under the additional assumption that the sectional curvatures in the Ricci flow are $O(t^{-1})$ as $t \to \infty$.

The paper [51] can be divided into four main parts.

Sections I.1-I.6 construct certain entropy-type functionals $\mathcal{F}$ and $\mathcal{W}$ that are monotonic under Ricci flow. The functional $\mathcal{W}$ is used to prove a no local collapsing theorem.

Sections I.7-I.10 introduce and apply another monotonic quantity, the reduced volume $\tilde{V}$. It is also used to prove a no local collapsing theorem. The construction of $\tilde{V}$ uses a modified notion of a geodesic, called an $\mathcal{L}$-geodesic. For technical reasons the reduced volume $\tilde{V}$ seems to be easier to work with than the $\mathcal{W}$-functional, and is used in most of the sequel. A reader who wants to focus on the Poincaré Conjecture or the Geometrization Conjecture could in principle start with I.7.

Section I.11 is concerned with κ-solutions, meaning nonflat ancient solutions that are κ-noncollapsed at all scales (for some $\kappa > 0$) and have bounded nonnegative curvature operator on each time slice. In three dimensions, a blowup limit of a finite-time singularity will be a κ-solution.

Sections I.12-I.13 are about three-dimensional Ricci flow solutions. It is shown that high-scalar-curvature regions are modeled by rescalings of κ-solutions. A decomposition of the time-t manifold into "thick" and "thin" pieces is described. It is stated that as $t \to \infty$, the thick piece becomes more and more hyperbolic, with incompressible cuspidal tori, and the thin piece is a graph manifold. More details of these assertions appear in [52], which also deals with the necessary modifications if the solution has singularities.

We now describe each of these four parts in a bit more detail.

4.1. **I.1-I.6.** In these sections M is assumed to be a closed n-dimensional manifold.

A functional $F(g)$ of Riemannian metrics g is said to be monotonic under Ricci flow if $F(g(t))$ is nondecreasing in t whenever $g(\cdot)$ is a Ricci flow solution. Monotonic quantities are an important tool for understanding Ricci flow. One wants to have useful monotonic quantities, in particular with a characterization of the Ricci flows for which $F(g(t))$ is constant in t.

Formally thinking of Ricci flow as a flow on the space of metrics, one way to get a monotonic quantity would be if the Ricci flow were the gradient flow of a functional F. In Sections I.1-I.2, a functional $\mathcal{F}$ is introduced whose gradient flow is not quite Ricci flow, but only differs from the Ricci flow by the action of diffeomorphisms. (If one formally considers the Ricci flow as a flow on the space of metrics modulo diffeomorphisms then it turns out to be the gradient flow of a functional λ_1.) The functional $\mathcal{F}$ actually depends on a Riemannian metric g and a function f. If $g(\cdot)$ satisfies the Ricci flow equation and $e^{-f(\cdot)}$ satisfies a conjugate or "backward" heat equation, in terms of $g(\cdot)$, then $\mathcal{F}(g(t), f(t))$ is nondecreasing in t. Furthermore, it is constant in t if and only if $g(\cdot)$ is a gradient steady soliton with associated function $f(\cdot)$. Minimizing $\mathcal{F}(g, f)$ over all functions f with $\int_M e^{-f} \, dV = 1$ gives the monotonic quantity $\lambda_1(g)$, which turns out to be the lowest eigenvalue of $- 4 \triangle + R$.

In Section I.3 a modified "entropy" functional $\mathcal{W}(g, f, \tau)$ is introduced. It is nondecreasing in t provided that $g(\cdot)$ is a Ricci flow, $\tau = t_0 - t$ and $(4\pi\tau)^{-\frac{n}{2}} e^{-f}$ satisfies the conjugate heat equation. The functional $\mathcal{W}$ is constant on a Ricci flow if and only if the flow is a gradient shrinking soliton that terminates at time t_0. Because of this, $\mathcal{W}$ is more suitable than $\mathcal{F}$ when one wants information that is localized in spacetime.

In Section I.4 the entropy functional $\mathcal{W}$ is used to prove a no local collapsing theorem. The statement is that if $g(\cdot)$ is a given Ricci flow on a finite time interval $[0, T)$ then for any (scale) $\rho > 0$, there is a number $\kappa > 0$ so that if $B_t(x, r)$ is a time-t ball with radius r less than ρ, on which $|\operatorname{Rm}| \le \frac{1}{r^2}$, then $\operatorname{vol}(B_t(x, r)) \ge \kappa r^n$. The method of proof is to show that if $r^{-n} \operatorname{vol}(B_t(x, r))$ is very small then the evaluation of $\mathcal{W}$ at time t is very negative, which contradicts the monotonicity of $\mathcal{W}$.

The significance of a no local collapsing theorem is that it allows one to use Hamilton's compactness theorem to construct blowup limits of finite time singularities, and more generally to understand high curvature regions.

Section I.5 and I.6 are not needed in the sequel. Section I.5 gives some thermodynamic-like equations in which $\mathcal{W}$ appears as an entropy. Section I.6 motivates the construction of the reduced volume of Section I.7.

4.2. **I.7-I.10.** A new monotonic quantity, the reduced volume $\tilde{V}$, is introduced in I.7. It is defined in terms of so-called $\mathcal{L}$-geodesics. Let (p, t_0) be a fixed spacetime point. Define backward time by $\tau = t_0 - t$. Given a curve $\gamma(\tau)$ in M defined for $0 \le \tau \le \overline{\tau}$ (i.e. going backward in real time) with $\gamma(0) = p$, its $\mathcal{L}$-length is

$$\mathcal{L}(\gamma) = \int_0^{\overline{\tau}} \sqrt{\tau} \, \left(|\dot{\gamma}(\tau)|^2 + R(\gamma(\tau), t_0 - \tau) \right) d\tau. \tag{4.1}$$

One can compute the first and second variations of $\mathcal{L}$, in analogy to what is done in Riemannian geometry.

Let $L(q, \overline{\tau})$ be the infimum of $\mathcal{L}(\gamma)$ over curves γ with $\gamma(0) = p$ and $\gamma(\overline{\tau}) = q$. Put $l(q, \overline{\tau}) = \frac{L(q,\overline{\tau})}{2\sqrt{\overline{\tau}}}$. The reduced volume is defined by $\tilde{V}(\overline{\tau}) = \int_M \overline{\tau}^{-\frac{n}{2}} e^{-l(q,\overline{\tau})} \operatorname{dvol}(q)$. The remarkable fact is that if $g(\cdot)$ is a Ricci flow solution then $\tilde{V}$ is nonincreasing in $\overline{\tau}$, i.e. nondecreasing in real time t. Furthermore, it is constant if and only if $g(\cdot)$ is a gradient shrinking soliton that terminates at time t_0. The proof of monotonicity uses a subtle cancellation between the $\overline{\tau}$-derivative of $l(\gamma(\overline{\tau}), \overline{\tau})$ along an $\mathcal{L}$-geodesic and the Jacobian of the so-called $\mathcal{L}$-exponential map.

Using a differential inequality, it is shown that for each $\overline{\tau}$ there is some point $q(\overline{\tau}) \in M$ so that $l(q(\overline{\tau}), \overline{\tau}) \le \frac{n}{2}$. This is then used to prove a no local collapsing theorem : Given a Ricci flow solution $g(\cdot)$ defined on a finite time interval $[0, t_0]$ and a scale $\rho > 0$ there is a number $\kappa > 0$ with the following property. For $r < \rho$, suppose that $|\operatorname{Rm}| \le \frac{1}{r^2}$ on the "parabolic" ball $\{(x, t) : \operatorname{dist}_{t_0}(x, p) \le r, \ t_0 - r^2 \le t \le t_0\}$. Then $\operatorname{vol}(B_{t_0}(p, r)) \ge \kappa r^n$. The number κ can be chosen to depend on ρ, n, t_0 and bounds on the geometry of the initial metric $g(0)$.

The hypotheses of the no local collapsing theorem proved using $\tilde{V}$ are more stringent than those of the no local collapsing theorem proved using $\mathcal{W}$, but the consequences turn out to

be the same. The no local collapsing theorem is used extensively in Sections I.11 and I.12 when extracting a convergent subsequence from a sequence of Ricci flow solutions.

Theorem I.8.2 of Section I.8 says that under appropriate assumptions, local κ-noncollapsing extends forwards in time to larger distances. This will be used in I.12 to analyze long-time behavior. The statement is that for any $A < \infty$ there is a $\kappa = \kappa(A) > 0$ with the following property. Given $r_0 > 0$ and $x_0 \in M$, suppose that $g(\cdot)$ is defined for $t \in [0, r_0^2]$, with $|\operatorname{Rm}(x,t)| \leq r_0^{-2}$ for all (x,t) satisfying $\operatorname{dist}_0(x, x_0) < r_0$, and the volume of the time-zero ball $B_0(x_0, r_0)$ is at least $A^{-1} r_0^n$. Then the metric $g(t)$ cannot be κ-collapsed on scales less than r_0 at a point (x, r_0^2) with $\operatorname{dist}_{r_0^2}(x, x_0) \leq A r_0$.

A localized version of the $\mathcal{W}$-functional appears in I.9. Section I.10, which is not needed for the sequel but is of independent interest, shows if a point in a time slice lies in a ball with quantitatively bounded geometry then at nearby later times, the curvature at the point is quantitatively bounded. That is, there is a damping effect with regard to exterior high-curvature regions. The "bounded geometry" assumptions on the initial ball are a lower bound on its scalar curvature and an assumption that the isoperimetric constants of subballs are close to the Euclidean value.

4.3. **I.11.** Section I.11 contains an analysis of κ-solutions. As mentioned before, in three dimensions κ-solutions arise as blowup limits of finite-time singularities and, more generally, as limits of rescalings of high-scalar-curvature regions.

In addition to the no local collapsing theorem, some of the tools used to analyze κ-solutions are Hamilton's Harnack inequality for Ricci flows with nonnegative curvature operator, and the comparison geometry of nonnegatively curved manifolds.

The first result is that any time slice of a κ-solution has vanishing asymptotic volume ratio $\lim_{r \to \infty} r^{-n} \operatorname{vol}(B_t(p,r))$. This apparently technical result is used to show that if a κ-solution $(M, g(\cdot))$ has scalar curvature normalized to equal one at some spacetime point (p,t) then there is an *a priori* upper bound on the scalar curvature $R(q,t)$ at other points q in terms of $\operatorname{dist}_t(p,q)$. Using the curvature bound, it is shown that a sequence $\{(M_i, p_i, g_i(\cdot))\}_{i=1}^\infty$ of pointed n-dimensional κ-solutions, normalized so that $R(p_i, 0) = 1$ for each i, has a convergent subsequence whose limit satisfies all of the requirements to be a κ-solution, except possibly the condition of having bounded sectional curvature on each time slice.

In three dimensions this statement is improved by showing that the sectional curvature will be bounded on each compact time interval, so the space of pointed 3-dimensional κ-solutions $(M, p, g(\cdot))$ with $R(p, 0) = 1$ is in fact compact. This is used to draw conclusions about the global geometry of 3-dimensional κ-solutions.

If M is a compact 3-dimensional κ-solution then Hamilton's theorem about compact 3-manifolds with nonnegative Ricci curvature implies that M is diffeomorphic to a spherical space form. If M is noncompact then assuming that M is oriented, it follows easily that M is diffeomorphic to $\mathbb{R}^3$, or isometric to the round shrinking cylinder $\mathbb{R} \times S^2$ or its $\mathbb{Z}_2$-quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^2$.

In the noncompact case it is shown that after rescaling, each time slice is neck-like at infinity. More precisely, considering a given time-t slice, for each $\epsilon > 0$ there is a compact

subset $M_\epsilon \subset M$ so that if $y \notin M_\epsilon$ then the pointed manifold $(M, y, R(y,t)g(t))$ is ϵ-close to the standard cylinder $\left[-\frac{1}{\epsilon}, \frac{1}{\epsilon}\right] \times S^2$ of scalar curvature one.

4.4. **I.12-I.13.** Sections I.12 and I.13 deal with three-dimensional Ricci flows.

Theorem I.12.1 uses the results of Section I.11 to model the high-scalar-curvature regions of a Ricci flow. Let us assume a pinching condition of the form $\mathrm{Rm} \geq -\Phi(R)$ for an appropriate function Φ with $\lim_{s\to\infty} \frac{\Phi(s)}{s} = 0$. (This will eventually follow from Hamilton-Ivey pinching, cf. Appendix B.) Theorem I.12.1 says that given numbers $\epsilon, \kappa > 0$, one can find $r_0 > 0$ with the following property. Suppose that $g(\cdot)$ is a Ricci flow solution defined on some time interval $[0,T]$ that satisfies the pinching condition and is κ-noncollapsed at scales less than one. Then for any point (x_0, t_0) with $t_0 \geq 1$ and $Q = R(x_0, t_0) \geq r_0^{-2}$, after scaling by the factor Q, the solution in the region $\{(x,t) : \mathrm{dist}^2_{t_0}(x, x_0) \leq (\epsilon Q)^{-1}, t_0 - (\epsilon Q)^{-1} \leq t \leq t_0\}$ is ϵ-close to the corresponding subset of a κ-solution.

Theorem I.12.1 says in particular that near a first singularity, the geometry is modeled by a κ-solution, for some κ. This fact is used in [52]. Although Theorem I.12.1 is not used directly in [51], its method of proof is used in Theorem I.12.2.

The method of proof of Theorem I.12.1 is by contradiction. If it were not true then there would be a sequence $r_0^{(i)} \to 0$ and a sequence $(M_i, g_i(\cdot))$ of Ricci flow solutions that satisfy the assumptions, each with a spacetime point $\left(x_0^{(i)}, t_0^{(i)}\right)$ that does not satisfy the conclusion. To consider first a special case, suppose that each point $\left(x_0^{(i)}, t_0^{(i)}\right)$ is the first point at which a certain curvature threshold R_i is achieved, i.e. $R(y,t) \leq R\left(x_0^{(i)}, t_0^{(i)}\right)$ for each $y \in M_i$ and $t \in \left[0, t_0^{(i)}\right]$. Then after rescaling the Ricci flow $g_i(\cdot)$ by $Q_i = R\left(x_0^{(i)}, t_0^{(i)}\right)$ and shifting the time parameter, one has the curvature bounds on the time interval $[-Q_i t_0^{(i)}, 0]$ that form part of the hypotheses of Hamilton's compactness theorem. Furthermore, the no local collapsing theorem gives the lower injectivity radius bound needed to apply Hamilton's theorem and take a convergent subsequence of the pointed rescaled solutions. The limit will be a κ-solution, giving the contradiction.

In the general case, one effectively proceeds by induction on the size of the scalar curvature. By modifying the choice of points $\left(x_0^{(i)}, t_0^{(i)}\right)$, one can assume that the conclusion of the theorem holds for all of the points (y,t) in a large spacetime neighborhood of $\left(x_0^{(i)}, t_0^{(i)}\right)$ that have $R(y,t) > 2Q_i$. One then shows that one has the curvature bounds needed to form the time-zero slice of the putative κ-solution. One shows that this "time-zero" metric can be extended backward in time to form a κ-solution, thereby giving the contradiction.

The rest of Section I.12 begins the analysis of the long-time behaviour of a nonsingular 3-dimensional Ricci flow. There are two main results, Theorems I.12.2 and I.12.3. They extend curvature bounds forward and backward in time, respectively.

Theorem I.12.2 roughly says that if one has $|\mathrm{Rm}| \leq r_0^{-2}$ on a spacetime region of spatial size r_0 and temporal size r_0^2, and if one has a lower bound on the volume of the initial time face of the region, then one gets scalar curvature bounds on much larger spatial balls at

the final time. More precisely, for any $A < \infty$ there are numbers $K = K(A) < \infty$ and $\rho = \rho(A) < \infty$ so that with the hypotheses and notation of Theorem I.8.2, if in addition $r_0^2\, \Phi(r_0^{-2}) < \rho$ then $R(x, r_0^2) \le K r_0^{-2}$ for points x lying in the ball of radius Ar_0 around x_0 at time r_0^2.

Theorem I.12.3 says that if one has a lower bound on volume and sectional curvature on a ball at a certain time then one obtains an upper scalar curvature bound on a smaller ball at an earlier time. More precisely, given $w > 0$ there exist $\tau = \tau(w) > 0$, $\rho = \rho(w) > 0$ and $K = K(w) < \infty$ with the following property. Suppose that a ball $B(x_0, r_0)$ at time t_0 has volume bounded below by wr_0^3 and sectional curvature bounded below by $-\, r_0^{-2}$. Then $R(x,t) < K r_0^{-2}$ for $t \in [t_0 - \tau r_0^2, t_0]$ and $\mathrm{dist}_t(x, x_0) < \frac{1}{4} r_0$, provided that $\phi(r_0^{-2}) < \rho$.

Applying a back-and-forth argument using Theorems I.12.2 and I.12.3, along with the pinching condition, one concludes, roughly speaking, that if a metric ball of small radius r has infimal sectional curvature exactly equal to $-\, r^{-2}$ then the ball has a small volume compared to r^3. Such a ball can be said to be locally volume-collapsed with respect to a lower sectional curvature bound.

Section I.13 defines the thick-thin decomposition of a large-time slice of a nonsingular Ricci flow and shows the geometrization. Rescaling the metric to $\widehat{g}(t) = t^{-1}\, g(t)$, there is a universal function Φ so that for large t, the metric $\widehat{g}(t)$ satisfies the Φ-pinching condition. In terms of the original unscaled metric, given $x \in M$ let $\widehat{r}(x,t) > 0$ be the unique number such that $\inf \mathrm{Rm}\big|_{B_t(x,\widehat{r})} = -\, \widehat{r}^{-2}$.

Given $w > 0$, define the w-thin part $M_{thin}(w,t)$ of the time-t slice to be the points $x \in M$ so that $\mathrm{vol}(B_t(x, \widehat{r}(x,t))) < w\, \widehat{r}(x,t)^3$. That is, a point in $M_{thin}(w,t)$ lies in a ball that is locally volume-collapsed with respect to a lower sectional curvature bound. Put $M_{thick}(w,t) = M - M_{thin}(w,t)$. One shows that for large t, the subset $M_{thick}(w,t)$ has bounded geometry in the sense that there are numbers $\overline{\rho} = \overline{\rho}(w) > 0$ and $K = K(w) < \infty$ so that $|\,\mathrm{Rm}\,| \le K t^{-1}$ on $B(x, \overline{\rho}\sqrt{t})$ and $\mathrm{vol}(B(x, \overline{\rho}\sqrt{t})) \ge \frac{1}{10}\, w\, (\overline{\rho}\sqrt{t}))^3$, whenever $x \in M_{thick}(w,t)$.

Invoking arguments of Hamilton (that are written out in more detail in [52]) one can take a sequence $t \to \infty$ and $w \to 0$ so that $M_{thick}(w,t)$ converges to a complete finite-volume manifold with constant sectional curvature $-\, \frac{1}{4}$, whose cuspidal tori are incompressible in M. On the other hand, a result from Riemannian geometry implies that for large t and small w, $M_{thin}(w,t)$ is homeomorphic to a graph manifold; again a more precise statement appears in [52]. The conclusion is that M satisfies the geometrization conjecture. Again, one is assuming in I.13 that the Ricci flow is nonsingular for all times.

5. I.1.1. The $\mathcal{F}$-functional and its monotonicity

The goal of this section is to show that in an appropriate sense, Ricci flow is a gradient flow on the space of metrics. We introduce the entropy functional $\mathcal{F}$. We compute its formal variation and show that the corresponding gradient flow is a modified Ricci flow.

In Sections 5 through 14 of these notes, M is a closed manifold. We will use the Einstein summation convention freely. We also follow Perelman's convention that a condition like $a > 0$ means that a should be considered to be a small parameter, while a condition like

$A < \infty$ means that A should be considered to be a large parameter. This convention is only for pedagogical purposes and may be ignored by a logically minded reader.

Let $\mathcal{M}$ denote the space of smooth Riemannian metrics g on M. We think of $\mathcal{M}$ formally as an infinite-dimensional manifold. The tangent space $T_g\mathcal{M}$ consists of the symmetric covariant 2-tensors v_{ij} on M. Similarly, $C^\infty(M)$ is an infinite-dimensional manifold with $T_f C^\infty(M) = C^\infty(M)$. The diffeomorphism group $\mathrm{Diff}(M)$ acts on $\mathcal{M}$ and $C^\infty(M)$ by pullback.

Let dV denote the Riemannian volume density associated to a metric g. We use the convention that $\triangle = \text{div grad}$.

Definition 5.1. The $\mathcal{F}$-functional $\mathcal{F} : \mathcal{M} \times C^\infty(M) \to \mathbb{R}$ is given by

$$\mathcal{F}(g, f) = \int_M \left(R + |\nabla f|^2\right) e^{-f} \, dV. \tag{5.2}$$

Given $v_{ij} \in T_g\mathcal{M}$ and $h \in T_f C^\infty(M)$, the evaluation of the differential $d\mathcal{F}$ on (v_{ij}, h) is written as $\delta\mathcal{F}(v_{ij}, h)$. Put $v = g^{ij} v_{ij}$.

Proposition 5.3. *(cf. I.1.1) We have*

$$\delta\mathcal{F}(v_{ij}, h) = \int_M e^{-f} \left[- v_{ij}(R_{ij} + \nabla_i\nabla_j f) + \left(\frac{v}{2} - h\right) (2\triangle f - |\nabla f|^2 + R)\right] dV. \tag{5.4}$$

Proof. From a standard formula,

$$\delta R = - \triangle v + \nabla_i\nabla_j v_{ij} - R_{ij} v_{ij}. \tag{5.5}$$

As

$$|\nabla f|^2 = g^{ij} \nabla_i f \nabla_j f, \tag{5.6}$$

we have

$$\delta |\nabla f|^2 = - v^{ij} \nabla_i f \nabla_j f + 2 \langle \nabla f, \nabla h \rangle. \tag{5.7}$$

As $dV = \sqrt{\det(g)} \, dx_1 \dots dx_n$, we have $\delta(dV) = \frac{v}{2} \, dV$, so

$$\delta \left(e^{-f} \, dV\right) = \left(\frac{v}{2} - h\right) e^{-f} \, dV. \tag{5.8}$$

Putting this together gives

$$\begin{aligned}\delta\mathcal{F} = \int_M e^{-f} \, [- \triangle v + \nabla_i\nabla_j v_{ij} - R_{ij} v_{ij} - v_{ij} \nabla_i f \nabla_j f + \\ 2 \langle \nabla f, \nabla h \rangle + (R + |\nabla f|^2) \left(\frac{v}{2} - h\right)\Big] \, dV.\end{aligned} \tag{5.9}$$

The goal now is to rewrite the right-hand side of (5.9) so that v_{ij} and h appear algebraically, i.e. without derivatives. As

$$\triangle e^{-f} = (|\nabla f|^2 - \triangle f) \, e^{-f}, \tag{5.10}$$

we have

$$\int_M e^{-f} \, [- \triangle v] \, dV = - \int_M (\triangle e^{-f}) \, v \, dV = \int_M e^{-f} \, (\triangle f - |\nabla f|^2) \, v \, dV. \tag{5.11}$$

Next,

$$\int_M e^{-f}\nabla_i\nabla_j v_{ij}\,dV \;=\; \int_M (\nabla_i\nabla_j e^{-f})\,v_{ij}\,dV \;=\; -\int_M \nabla_i(e^{-f}\nabla_j f)\,v_{ij}\,dV \tag{5.12}$$
$$= \int_M e^{-f}(\nabla_i f\nabla_j f \;-\; \nabla_i\nabla_j f)\,v_{ij}\,dV.$$

Finally,

$$2\int_M e^{-f}\langle\nabla f,\nabla h\rangle\,dV \;=\; -\,2\int_M\langle\nabla e^{-f},\nabla h\rangle\,dV \;=\; 2\int_M(\triangle e^{-f})\,h\,dV \tag{5.13}$$
$$= 2\int_M e^{-f}\,(|\nabla f|^2 \;-\; \triangle f)\,h\,dV.$$

Then

$$\delta\mathcal{F} \;=\; \int_M e^{-f}\,\Big[\Big(\frac{v}{2} - h\Big)(2\triangle f \;-\; 2|\nabla f|^2) \;-\; v_{ij}(R_{ij} \;+\; \nabla_i\nabla_j f) \tag{5.14}$$
$$+\Big(\frac{v}{2} - h\Big)\,(R \;+\; |\nabla f|^2)\Big]\,dV$$
$$= \int_M e^{-f}\,\Big[-\,v_{ij}(R_{ij} \;+\; \nabla_i\nabla_j f) \;+\; \Big(\frac{v}{2} - h\Big)\,(2\triangle f \;-\; |\nabla f|^2 \;+\; R)\Big]\,dV.$$

This proves the proposition. □

We would like to get rid of the $\left(\frac{v}{2} - h\right)\,(2\triangle f \,-\, |\nabla f|^2 \,+\, R)$ term in (5.14). We can do this by restricting our variations so that $\frac{v}{2} - h \,=\, 0$. From (5.8), this amounts to assuming that assuming $e^{-f}\,dV$ is fixed. We now fix a smooth measure dm on M and relate f to g by requiring that $e^{-f}\,dV \,=\, dm$. Equivalently, we define a section $s : \mathcal{M} \to \mathcal{M}\times C^\infty(M)$ by $s(g) \,=\, \left(g, \ln\left(\frac{dV}{dm}\right)\right)$. Then the composition $\mathcal{F}^m \,=\, \mathcal{F}\circ s$ is a function on $\mathcal{M}$ and its differential is given by

$$d\mathcal{F}^m(v_{ij}) \;=\; \int_M e^{-f}\,[-\,v_{ij}(R_{ij} \;+\; \nabla_i\nabla_j f)]\,dV. \tag{5.15}$$

Defining a formal Riemannian metric on $\mathcal{M}$ by

$$\langle v_{ij}, v_{ij}\rangle_g \;=\; \frac{1}{2}\int_M v^{ij}\,v_{ij}\,dm, \tag{5.16}$$

the gradient flow of $\mathcal{F}^m$ on $\mathcal{M}$ is given by

$$(g_{ij})_t \;=\; -2\,(R_{ij} \;+\; \nabla_i\nabla_j f). \tag{5.17}$$

The induced flow equation for f is

$$f_t \;=\; \frac{\partial}{\partial t}\ln\left(\frac{dV}{dm}\right) \;=\; \frac{1}{2}\,g^{ij}\,(g_{ij})_t \;=\; -\,\triangle f \;-\; R. \tag{5.18}$$

As with any gradient flow, the function $\mathcal{F}^m$ is nondecreasing along the flow line with its derivative being given by the length squared of the gradient, i.e.

$$\mathcal{F}^m_t \;=\; 2\int_M |R_{ij} \;+\; \nabla_i\nabla_j f|^2\,dm, \tag{5.19}$$

as follows from (5.14) and (5.17)

We now perform time-dependent diffeomorphisms to transform (5.17) into the Ricci flow equation. If $V(t)$ is the time-dependent generating vector field of the diffeomorphisms then the new equations for g and f become

$$\begin{aligned}(g_{ij})_t \;&=\; -2\left(R_{ij} \,+\, \nabla_i\nabla_j f\right) \,+\, \mathcal{L}_V g,\\ f_t \;&=\; -\,\triangle f \,-\, R \,+\, \mathcal{L}_V f.\end{aligned} \tag{5.20}$$

Taking $V \;=\; \nabla f$ gives

$$\begin{aligned}(g_{ij})_t \;&=\; -2\,R_{ij},\\ f_t \;&=\; -\,\triangle f \,-\, R \,+\, |\nabla f|^2.\end{aligned} \tag{5.21}$$

As $\mathcal{F}(g,f)$ is unchanged by a simultaneous pullback of g and f, and the right-hand side of (5.19) is also unchanged under a simultaneous pullback, it follows that under the new evolution equations (5.21) we still have

$$\frac{d}{dt}\,\mathcal{F}(g(t),f(t)) \;=\; 2\int_M |R_{ij} \,+\, \nabla_i\nabla_j f|^2\, e^{-f}\, dV \tag{5.22}$$

(This can also be checked directly).

Because of the diffeomorphisms that we applied, g and f are no longer related by $e^{-f}\,dV = dm$. We do have that $\int_M e^{-f}\,dV$ is constant in t, as $e^{-f}\,dV$ is related to dm by a diffeomorphism.

The relation between g and f is as follows: we solve (or assume that we have a solution for) the first equation in (5.21), with some initial metric. Then given the solution $g(t)$, we require that f satisfy the second equation in (5.21) (which is in terms of $g(t)$).

The second equation in (5.21) can be written as

$$\frac{\partial}{\partial t}\,e^{-f} \;=\; -\,\triangle e^{-f} \,+\, R\,e^{-f}. \tag{5.23}$$

As this is a backward heat equation, we cannot solve for f forward in time starting with an arbitrary smooth function. Instead, (5.21) will be applied by starting with a solution for $(g_{ij})_t \;=\; -2\,R_{ij}$ on some time interval $[t_1,t_2]$ and then solving (5.23) backwards in time on $[t_1,t_2]$ (which can always be done) starting with some initial $f(t_2)$. Having done this, the solution $(g(t),f(t))$ on $[t_1,t_2]$ will satisfy (5.22).

6. Basic example for I.1

In this section we compute $\mathcal{F}$ in a Euclidean example.

Consider $\mathbb{R}^n$ with the standard metric, constant in time. Fix $t_0 \,>\, 0$. Put $\tau \,=\, t_0 - t$ and

$$f(t,x) \;=\; \frac{|x|^2}{4\tau} \,+\, \frac{n}{2}\ln(4\pi\tau), \tag{6.1}$$

so

$$e^{-f} \;=\; (4\pi\tau)^{-n/2}\, e^{-\frac{|x|^2}{4\tau}}. \tag{6.2}$$

This is the standard heat kernel when considered for τ going from 0 to t_0, i.e. for t going from t_0 to 0. One can check that (g, f) solves (5.21). As

$$\int_{\mathbb{R}^n} e^{-\frac{|x|^2}{4\tau}} \, dV \;=\; (4\pi\tau)^{n/2}, \tag{6.3}$$

f is properly normalized. Then $\nabla f = \frac{\mathrm{x}}{2\tau}$ and $|\nabla f|^2 = \frac{|x|^2}{4\tau^2}$. Differentiating (6.3) with respect to τ gives

$$\int_{\mathbb{R}^n} \frac{|x|^2}{4\tau^2} \, e^{-\frac{|x|^2}{4\tau}} \, dV \;=\; (4\pi\tau)^{n/2} \, \frac{n}{2\tau}, \tag{6.4}$$

so

$$\int_{\mathbb{R}^n} |\nabla f|^2 \, e^{-f} \, dV \;=\; \frac{n}{2\tau}. \tag{6.5}$$

Then $\mathcal{F}(t) = \frac{n}{2\tau} = \frac{n}{2(t_0 - t)}$. In particular, this is nondecreasing as a function of $t \in [0, t_0)$.

7. I.2.2. The λ-invariant and its applications

In this section we define $\lambda(g)$ and show that it is nondecreasing under Ricci flow. We use this to show that a steady breather on a compact manifold is a gradient steady soliton.

Proposition 7.1. *Given a metric g, there is a unique minimizer $\overline{f}$ of $\mathcal{F}(g, f)$ under the constraint $\int_M e^{-f} \, dV = 1$.*

Proof. Write

$$\mathcal{F} \;=\; \int_M \left(R e^{-f} \;+\; 4 |\nabla e^{-f/2}|^2 \right) \, dV. \tag{7.2}$$

Putting $\Phi = e^{-f/2}$,

$$\mathcal{F} \;=\; \int_M \left(4 |\nabla \Phi|^2 \;+\; R\,\Phi^2 \right) \, dV \;=\; \int_M \Phi \left(-\,4 \triangle \Phi \;+\; R \Phi \right) \, dV. \tag{7.3}$$

The constraint equation becomes $\int_M \Phi^2 dV = 1$. Then λ is the smallest eigenvalue of $-4\triangle + R$ and $e^{-\overline{f}/2}$ is a corresponding normalized eigenvector. As the operator is a Schrödinger operator, there is a unique normalized positive eigenvector [55, Chapter XIII.12]. □

Definition 7.4. The λ-functional is given by $\lambda(g) = \mathcal{F}(g, \overline{f})$.

If $g(t)$ is a smooth family of metrics then it follows from eigenvalue perturbation theory that $\lambda(g(t))$ and $\overline{f}(t)$ are smooth in t [55, Chapter XII].

Proposition 7.5. *(cf. I.2.2) If $g(\cdot)$ is a Ricci flow solution then $\lambda(g(t))$ is nondecreasing in t.*

Proof. Consider a time interval $[t_1, t_2]$, and the minimizer $\overline{f}(t_2)$. In particular, $\lambda(t_2) = \mathcal{F}(g(t_2), \overline{f}(t_2))$. Put $u(t_2) = e^{-\overline{f}(t_2)}$. Solve the backward heat equation

$$\frac{\partial u}{\partial t} \;=\; -\,\triangle u \;+\; R u \tag{7.6}$$

backward on $[t_1, t_2]$.

We claim that $u(x', t') > 0$ for all $x' \in M$ and $t' \in [t_1, t_2]$. To see this, we take $t' \in [t_1, t_2)$, and let h be the solution to the forward heat equation $\frac{\partial h}{\partial t} = \triangle h$ on $(t', t_2]$ with $\lim_{t \to t'} h(t) = \delta_{x'}$. We have

$$\frac{d}{dt} \int_M u(t)\, h(t)\, dV = \int_M \left[(\partial_t u + \triangle u - Ru)\, v + u\, (\partial_t h - \triangle h) \right] dV = 0. \tag{7.7}$$

One knows that $h(t) > 0$ for all $t \in (t', t_2]$. Then

$$u(x', t') = \int_M u(x, t')\, \delta_{x'}(x)\, dV(x) = \lim_{t \to t'} \int_M u(t)\, h(t)\, dV = \int_M u(t_2)\, h(t_2)\, dV > 0. \tag{7.8}$$

For $t \in [t_1, t_2]$, define $f(t)$ by $u(t) = e^{-f(t)}$. By (5.22), $\mathcal{F}(g(t_1), f(t_1)) \le \mathcal{F}(g(t_2), f(t_2))$. By the definition of λ, $\lambda(t_1) = \mathcal{F}(g(t_1), \overline{f}(t_1)) \le \mathcal{F}(g(t_1), f(t_1))$. (We are using the fact that $\int_M e^{-f(t_1)}\, dV(t_1) = \int_M e^{-f(t_2)}\, dV(t_2) = 1$.) Thus $\lambda(t_1) \le \lambda(t_2)$. □

Definition 7.9. A *steady breather* is a Ricci flow solution on an interval $[t_1, t_2]$ that satisfies the equation $g(t_2) = \phi^* g(t_1)$ for some $\phi \in \mathrm{Diff}(M)$.

Steady soliton solutions are steady breathers.

Again, we are assuming that M is compact. The next result is not essential for the sequel, but gives a good illustration of how a monotonicity formula is used.

Proposition 7.10. *(cf. I.2.2) A steady breather is a gradient steady soliton.*

Proof. We have $\lambda(g(t_2)) = \lambda(\phi^* g(t_1)) = \lambda(g(t_1))$. Thus we have equality in Proposition 7.5. Tracing through the proof, $\mathcal{F}(g(t), f(t))$ must be constant in t. From (5.22), $R_{ij} + \nabla_i \nabla_j f = 0$. Then $R + \triangle f = 0$ and so (5.21) becomes (C.5). □

One can sharpen Proposition 7.5.

Lemma 7.11. *(cf. Proposition I.1.2)*

$$\frac{d\lambda}{dt} \ge \frac{2}{n} \lambda^2(t). \tag{7.12}$$

Proof. Given a time interval $[t_1, t_2]$, with the notation of the proof of Proposition 7.5 we have

$$\begin{aligned} \lambda(t_1) &\le \mathcal{F}(g(t_1), f(t_1)) = \mathcal{F}(g(t_2), f(t_2)) - 2 \int_{t_1}^{t_2} \int_M |R_{ij} + \nabla_i \nabla_j f|^2\, e^{-f}\, dV\, dt \\ &= \lambda(t_2) - 2 \int_{t_1}^{t_2} \int_M |R_{ij} + \nabla_i \nabla_j f|^2\, e^{-f}\, dV\, dt. \end{aligned} \tag{7.13}$$

Then

$$\left. \frac{d\lambda}{dt} \right|_{t = t_2} = \lim_{t_1 \to t_2^-} \frac{\lambda(t_2) - \lambda(t_1)}{t_2 - t_1} \ge 2 \int_M |R_{ij} + \nabla_i \nabla_j f|^2\, e^{-f}\, dV, \tag{7.14}$$

where the right-hand side is evaluated at time t_2.

Hence for all t,

$$\frac{d\lambda}{dt} \geq 2 \int_M |R_{ij} + \nabla_i \nabla_j \overline{f}|^2 \, e^{-\overline{f}} \, dV \tag{7.15}$$

and so

$$\frac{d\lambda}{dt} \geq \frac{2}{n} \int_M (R + \triangle \overline{f})^2 \, e^{-\overline{f}} \, dV. \tag{7.16}$$

From the Cauchy-Schwarz inequality and the fact that $\int_M e^{-\overline{f}} \, dV = 1$,

$$\left(\int_M (R + \triangle \overline{f}) \, e^{-\overline{f}} \, dV \right)^2 \leq \int_M (R + \triangle \overline{f})^2 \, e^{-\overline{f}} \, dV. \tag{7.17}$$

Finally, (5.10) gives

$$\int_M (R + \triangle \overline{f}) \, e^{-\overline{f}} \, dV = \int_M (R + |\nabla \overline{f}|^2) \, e^{-\overline{f}} \, dV = \mathcal{F}(g(t), \overline{f}(t)) = \lambda(t). \tag{7.18}$$

This proves the lemma. □

8. I.2.3. The rescaled λ-invariant

In this section we show the monotonicity of a scale-invariant version of λ. This will be used in Section 93. We then show that an expanding breather on a compact manifold is a gradient expanding soliton.

Put $\overline{\lambda}(g) = \lambda(g) V(g)^{\frac{2}{n}}$. As $\overline{\lambda}$ is scale-invariant, it is constant in t along a steady, shrinking or expanding soliton solution.

Proposition 8.1. *(cf. Claim of I.2.3) If $g(\cdot)$ is a Ricci flow solution and $\overline{\lambda}(g(t)) \leq 0$ for some t then $\frac{d}{dt}\overline{\lambda}(g(t)) \geq 0$.*

Proof. We have

$$\frac{d\overline{\lambda}}{dt} = \frac{d\lambda}{dt} V(t)^{\frac{2}{n}} - \frac{2}{n} V(t)^{\frac{2-n}{n}} \lambda(t) \int_M R \, dV. \tag{8.2}$$

From (7.15),

$$\frac{d\overline{\lambda}}{dt} \geq V(t)^{\frac{2}{n}} \left[2 \int_M |R_{ij} + \nabla_i \nabla_j \overline{f}|^2 \, e^{-\overline{f}} \, dV - \frac{2}{n} V(t)^{-1} \lambda(t) \int_M R \, dV \right]. \tag{8.3}$$

Using the spatially-constant function $\ln V(t)$ as a test function for $\mathcal{F}$ gives

$$\lambda(t) \leq V(t)^{-1} \int_M R \, dV. \tag{8.4}$$

The assumption that $\lambda(t) \leq 0$ gives

$$-\lambda(t)^2 \leq -V(t)^{-1} \lambda(t) \int_M R \, dV \tag{8.5}$$

and so

$$\frac{d\overline{\lambda}}{dt} \geq V(t)^{\frac{2}{n}} \left[2 \int_M |R_{ij} + \nabla_i \nabla_j \overline{f}|^2 \, e^{-\overline{f}} \, dV - \frac{2}{n} \lambda(t)^2 \right]. \tag{8.6}$$

Next,

$$|R_{ij} + \nabla_i \nabla_j \overline{f}|^2 \geq \left| R_{ij} + \nabla_i \nabla_j \overline{f} - \frac{1}{n} (R + \triangle \overline{f}) \, g_{ij} \right|^2 + \frac{1}{n} (R + \triangle \overline{f})^2. \tag{8.7}$$

Using (7.18), one obtains

(8.8)

$$\frac{d\overline{\lambda}}{dt} \geq 2 \, V(t)^{\frac{2}{n}} \left[\int_M \left| R_{ij} + \nabla_i \nabla_j \overline{f} - \frac{1}{n} (R + \triangle \overline{f}) \, g_{ij} \right|^2 e^{-\overline{f}} \, dV + \frac{1}{n} \int_M (R + \triangle \overline{f})^2 \, e^{-\overline{f}} \, dV \right.$$
$$\left. - \frac{1}{n} \left(\int_M (R + \triangle \overline{f}) \, e^{-\overline{f}} \, dV \right)^2 \right].$$

As $\int_M e^{-\overline{f}} \, dV = 1$, the Cauchy-Schwarz inequality implies that the right-hand side of (8.8) is nonnegative. □

Corollary 8.9. *If $\overline{\lambda}$ is a constant nonpositive number on an interval $[t_1, t_2]$ then the Ricci flow solution is a gradient soliton.*

Proof. From equation (8.8), we obtain that $R + \triangle \overline{f} = \alpha(t)$ for some function α that is spatially constant, and $R_{ij} + \nabla_i \nabla_j \overline{f} = \frac{\alpha(t)}{n} \, g_{ij}$. Thus g evolves by diffeomorphisms and dilations. After a shift of the time parameter, $\alpha(t)$ is proportionate to t cf. [22, Lemma 2.4]. This is the gradient soliton equation of Appendix C. □

Definition 8.10. An *expanding breather* is a Ricci flow solution on $[t_1, t_2]$ that satisfies $g(t_2) = c \, \phi^* g(t_1)$ for some $c > 1$ and $\phi \in \mathrm{Diff}(M)$.

Expanding soliton solutions are expanding breathers.

Again, we are assuming that M is compact.

Proposition 8.11. *An expanding breather is a gradient expanding soliton.*

Proof. First, $\overline{\lambda}(t_2) = \overline{\lambda}(t_1)$. As $V(t_2) > V(t_1)$, we must have $\frac{dV}{dt} > 0$ for some $t \in [t_1, t_2]$. From (8.4), $\frac{dV}{dt} = - \int_M R \, dV \leq - \lambda(t) \, V(t)$, so $\overline{\lambda}(t)$ must be negative for some $t \in [t_1, t_2]$. Proposition 8.1 implies that $\overline{\lambda}(t_1) < 0$. Then as $\overline{\lambda}(t_2) = \overline{\lambda}(t_1)$, it follows that $\overline{\lambda}$ is a negative constant on $[t_1, t_2]$. From Corollary 8.9, the solution is a gradient expanding soliton. □

9. I.2.4. Gradient steady solitons on compact manifolds

It was shown in Section 7 that a steady breather on a compact manifold is a gradient steady soliton. We now show that it is in fact Ricci flat. This was previously shown in [33, Theorem 20.1].

Proposition 9.1. *A gradient steady solution on a compact manifold is Ricci flat.*

Proof. As we are in the equality case of Proposition 7.5, the function $f(t)$ must be the minimizer of $\mathcal{F}(g(t), \cdot)$ for all t. That is,

$$(- 4\triangle + R) \, e^{-\frac{f}{2}} = \lambda \, e^{-\frac{f}{2}} \tag{9.2}$$

for all t, where λ is constant in t. Equivalently, $2\triangle f - |\nabla f|^2 + R = \lambda$. As $R + \triangle f = 0$, we have $\triangle f - |\nabla f|^2 = \lambda$. Then $\triangle e^{-f} = -\lambda e^{-f}$. Integrating gives $0 = \int_M \triangle e^{-f} dV = -\lambda \int_M e^{-f} dV$, so $\lambda = 0$. Then $0 = -\int_M e^{-f} \triangle e^{-f} \, dV = \int_M \left|\nabla e^{-f}\right|^2 dV$, so f is constant and g is Ricci flat. □

A similar argument shows that a gradient expanding soliton on a compact manifold comes from an Einstein metric with negative Ricci curvature.

10. Ricci flow as a gradient flow

We have shown in Section 5 that the modified Ricci flow is the gradient flow for the functional $\mathcal{F}^m$ on the space of metrics $\mathcal{M}$. One can ask if the unmodified Ricci flow is a gradient flow. This turns out to be true provided that one considers it as a flow on the space $\mathcal{M}/\operatorname{Diff}(M)$.

As mentioned in II.8, Ricci flow is the gradient flow for the function λ. More precisely, this statement is valid on $\mathcal{M}/\operatorname{Diff}(M)$, with the latter being equipped with an appropriate metric. To see this, we first consider λ as a function on the space of metrics $\mathcal{M}$. Here the formal Riemannian metric on $\mathcal{M}$ comes from saying that for $v_{ij} \in T_g\mathcal{M}$,

$$\langle v_{ij}, v_{ij} \rangle = \frac{1}{2} \int_M v^{ij} \, v_{ij} \, \Phi^2 \, dV(g), \tag{10.1}$$

where $\Phi = \Phi(g)$ is the unique normalized positive eigenvector corresponding to $\lambda(g)$.

Lemma 10.2. *The formal gradient flow of λ is*

$$\frac{\partial g_{ij}}{\partial t} = -2 \left(R_{ij} - 2 \nabla_i \nabla_j \ln \Phi\right). \tag{10.3}$$

Proof. We set

$$\lambda(g) = \inf_{f \in C^\infty(M) \,:\, \int_M e^{-f} \, dV = 1} \mathcal{F}(g, f). \tag{10.4}$$

To calculate the variation in λ due to a variation $\delta g_{ij} = v_{ij}$, we let $h = \delta f$ be the variation induced by letting f be the minimizer in (10.4). Then

$$0 = \delta \left(\int_M e^{-f} \, dV \right) = \int_M \left(\frac{v}{2} - h \right) e^{-f} \, dV. \tag{10.5}$$

Now equation (5.14) gives

$$\delta\lambda(v_{ij}) = \int_M e^{-f} \Big[- v_{ij}(R_{ij} + \nabla_i \nabla_j f) + \left(\frac{v}{2} - h \right) (2\triangle f - |\nabla f|^2 + R) \Big] \, dV. \tag{10.6}$$

As $\Phi = e^{-\frac{f}{2}}$ satisfies

$$-4 \triangle \Phi + R\Phi = \lambda\Phi, \tag{10.7}$$

it follows that

$$2 \triangle f - |\nabla f|^2 + R = \lambda. \tag{10.8}$$

Hence the last term in (10.6) vanishes, and (10.6) becomes

$$\delta\lambda(v_{ij}) \;=\; -\int_M e^{-f}\, v_{ij}(R_{ij} \;+\; \nabla_i\nabla_j f)\, dV. \tag{10.9}$$

The corresponding gradient flow is

$$\frac{\partial g_{ij}}{\partial t} \;=\; -\,2\,\left(R_{ij} \;+\; \nabla_i\nabla_j f\right) \;=\; -\,2\,\left(R_{ij} \;-\; 2\,\nabla_i\nabla_j \ln\Phi\right). \tag{10.10}$$

□

We note that it follows from (10.8) that

$$\nabla_j\left(\left(R_{ij} \;+\; \nabla_i\nabla_j f\right) e^{-f}\right) \;=\; 0. \tag{10.11}$$

This implies that the gradient vector field of λ is perpendicular to the infinitesimal diffeomorphisms at g, as one would expect.

In the sense of [10], the quotient space $\mathcal{M}/\operatorname{Diff}(M)$ is a stratified infinite-dimensional Riemannian manifold, with the strata corresponding to the possible isometry groups $\operatorname{Isom}(M, g)$. We give it the quotient Riemannian metric coming from (10.1). The modified Ricci flow (10.10) on $\mathcal{M}$ projects to a flow on $\mathcal{M}/\operatorname{Diff}(M)$ that coincides with the projection of the unmodified Ricci flow $\frac{dg}{dt} \;=\; -\,2\operatorname{Ric}$. The upshot is that the Ricci flow, as a flow on $\mathcal{M}/\operatorname{Diff}(M)$, is the gradient flow of λ, the latter now being considered as a function on $\mathcal{M}/\operatorname{Diff}(M)$.

One sees an intuitive explanation for Proposition 7.10. If a gradient flow on a finite-dimensional manifold has a periodic orbit then it must be a fixed-point. Applying this principle formally to the Ricci flow on $\mathcal{M}/\operatorname{Diff}(M)$, one infers that a steady breather only evolves by diffeomorphisms.

11. The $\mathcal{W}$-functional

Definition 11.1. The *$\mathcal{W}$-functional* $\mathcal{W} : \mathcal{M} \times C^\infty(M) \times \mathbb{R}^+ \to \mathbb{R}$ is given by

$$\mathcal{W}(g, f, \tau) \;=\; \int_M \left[\tau\left(|\nabla f|^2 + R\right) \;+\; f \;-\; n\right]\, (4\pi\tau)^{-\frac{n}{2}}\, e^{-f}\, dV. \tag{11.2}$$

The $\mathcal{W}$-functional is a scale-invariant variant of $\mathcal{F}$. It has the symmetries $\mathcal{W}(\phi^* g, \phi^* f, \tau) = \mathcal{W}(g, f, \tau)$ for $\phi \in \operatorname{Diff}(M)$, and $\mathcal{W}(cg, f, c\tau) \;=\; \mathcal{W}(g, f, \tau)$ for $c > 0$. Hence it is constant in $t \;=\; -\,\tau$ along a gradient shrinking soliton defined for $t \in (-\infty, 0)$, as in Appendix C. In this sense, $\mathcal{W}$ is constant on gradient shrinking solitons just as $\mathcal{F}$ is constant on gradient steady solitons.

As an example of a gradient shrinking soliton, consider $\mathbb{R}^n$ with the flat metric, constant in time $t \in (-\infty, 0)$. Put $\tau \;=\; -\,t$ and

$$f(t, x) \;=\; \frac{|x|^2}{4\tau}, \tag{11.3}$$

so

$$e^{-f} \;=\; e^{-\frac{|x|^2}{4\tau}}. \tag{11.4}$$

One can check that $(g(t), f(t), \tau(t))$ satisfies (12.12) and (12.13). Now

$$\tau(|\nabla f|^2 + R) + f - n = \tau \cdot \frac{|x|^2}{4\tau^2} + \frac{|x|^2}{4\tau} - n = \frac{|x|^2}{2\tau} - n. \tag{11.5}$$

It follows from (6.3) and (6.4) that $\mathcal{W}(t) = 0$ for all t.

12. I.3.1. Monotonicity of the $\mathcal{W}$-functional

In this section we compute the variation of $\mathcal{W}$, in analogy with the computation in Section 5 of the variation of $\mathcal{F}$. We then show that a shrinking breather on a compact manifold is a gradient shrinking soliton.

As in Section 5, we write $\delta g_{ij} = v_{ij}$ and $\delta f = h$. Put $\sigma = \delta\tau$.

Proposition 12.1. *We have*

(12.2)

$$\delta\mathcal{W}(v_{ij}, h, \sigma) = \int_M \Big[\sigma(R + |\nabla f|^2) - \tau v_{ij}(R_{ij} + \nabla_i\nabla_j f) + h + \Big[\tau(2\triangle f - |\nabla f|^2 + R) + f - n\Big]\left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right)\Big] (4\pi\tau)^{-n/2} e^{-f} \, dV.$$

Proof. One finds

$$\delta\left((4\pi\tau)^{-n/2} e^{-f} \, dV\right) = \left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right) (4\pi\tau)^{-n/2} e^{-f} \, dV. \tag{12.3}$$

Writing

$$\mathcal{W} = \int_M \Big[\tau(R + |\nabla f|^2) + f - n\Big] (4\pi\tau)^{-n/2} e^{-f} \, dV \tag{12.4}$$

we can use (5.14) to obtain

(12.5)

$$\delta\mathcal{W} = \int_M \Big[\sigma(R + |\nabla f|^2) + \tau\left(\frac{v}{2} - h\right)(2\triangle f - 2|\nabla f|^2) - \tau v_{ij}(R_{ij} + \nabla_i\nabla_j f) +$$

$$h + \Big[\tau(R + |\nabla f|^2) + f - n\Big]\left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right)\Big] (4\pi\tau)^{-n/2} e^{-f} \, dV. \tag{12.6}$$

Then (5.10) gives

(12.7)

$$\delta\mathcal{W} = \int_M \Big[\sigma(R + |\nabla f|^2) - \tau v_{ij}(R_{ij} + \nabla_i\nabla_j f) + h + \Big[\tau(2\triangle f - |\nabla f|^2 + R) + f - n\Big]\left(\frac{v}{2} - h - \frac{n\sigma}{2\tau}\right)\Big] (4\pi\tau)^{-n/2} e^{-f} \, dV.$$

This proves the proposition. □

We now fix a smooth measure dm on M with mass 1 and relate f to g and τ by requiring that $(4\pi\tau)^{-n/2} e^{-f} \, dV = dm$. Then $\frac{v}{2} - h - \frac{n\sigma}{2\tau} = 0$ and

$$\delta\mathcal{W} = \int_M \Big[\sigma(R + |\nabla f|^2) - \tau v_{ij}(R_{ij} + \nabla_i\nabla_j f) + h\Big] (4\pi\tau)^{-n/2} e^{-f} \, dV. \tag{12.8}$$

We now consider $\frac{d\mathcal{W}}{dt}$ when

$$
\begin{aligned}
(g_{ij})_t &= -2\,(R_{ij} + \nabla_i\nabla_j f),\\
f_t &= -\triangle f - R + \frac{n}{2\tau},\\
\tau_t &= -1.
\end{aligned}
\tag{12.9}
$$

To apply (12.8), we put

$$
\begin{aligned}
v_{ij} &= -2\,(R_{ij} + \nabla_i\nabla_j f),\\
h &= -\triangle f - R + \frac{n}{2\tau},\\
\sigma &= -1.
\end{aligned}
\tag{12.10}
$$

We do have $\frac{v}{2} - h - \frac{n\sigma}{2\tau} = 0$. Then from (12.8),

$$
\begin{aligned}
\frac{d\mathcal{W}}{dt} &= \int_M \big[-(R + |\nabla f|^2) + 2\,\tau|R_{ij} + \nabla_i\nabla_j f|^2\\
&\qquad - \triangle f - R + \frac{n}{2\tau}\Big]\,(4\pi\tau)^{-n/2}\,e^{-f}\,dV\\
&= \int_M \Big[-2(R + \triangle f) + 2\,\tau|R_{ij} + \nabla_i\nabla_j f|^2 + \frac{n}{2\tau}\Big]\,(4\pi\tau)^{-n/2}\,e^{-f}\,dV\\
&= \int_M 2\,\tau|R_{ij} + \nabla_i\nabla_j f - \frac{1}{2\tau}\,g_{ij}|^2\,(4\pi\tau)^{-n/2}\,e^{-f}\,dV.
\end{aligned}
\tag{12.11}
$$

Adding a Lie derivative to the right-hand side of (12.9) gives the new flow equations

$$
\begin{aligned}
(g_{ij})_t &= -2\,R_{ij},\\
f_t &= -\triangle f + |\nabla f|^2 - R + \frac{n}{2\tau},\\
\tau_t &= -1,
\end{aligned}
\tag{12.12}
$$

with (12.11) still holding. We no longer have $(4\pi\tau)^{-n/2}\,e^{-f}\,dV = dm$, but we do have

$$
\int_M (4\pi\tau)^{-n/2}\,e^{-f}\,dV = 1. \tag{12.13}
$$

We now want to look at the variational problem of minimizing $\mathcal{W}(g, f, \tau)$ under the constraint that $\int_M (4\pi\tau)^{-n/2}\,e^{-f}\,dV = 1$. We write

$$
\mu(g,\tau) = \inf_f\{\mathcal{W}(g, f, \tau) : \int_M (4\pi\tau)^{-n/2}\,e^{-f}\,dV = 1\}. \tag{12.14}
$$

Making the change of variable $\Phi = e^{-\frac{f}{2}}$, we are minimizing

$$
(4\pi\tau)^{-n/2}\int_M \left[\tau(4|\nabla\Phi|^2 + R\Phi^2) - 2\Phi^2\log\Phi - n\,\Phi^2\right]dV \tag{12.15}
$$

under the constraint $(4\pi\tau)^{-n/2}\int_M \Phi^2\,dV = 1$. From [56, Section 1] the infimum is finite and there is a positive continuous minimizer Φ. It will be a weak solution of the variational equation

$$
\tau(-4\triangle + R)\Phi = 2\Phi\log\Phi + (\mu(g,\tau) + n)\Phi. \tag{12.16}
$$

From elliptic theory, Φ is smooth. Then $f = -2\log\Phi$ is also smooth.

As in Section 7, it follows that $\mu(g(t), t_0 - t)$ is nondecreasing in t for a Ricci flow solution, where t_0 is any fixed number and $t < t_0$. If it is constant in t then the solution must be a gradient shrinking soliton that goes singular at time t_0.

Definition 12.17. A *shrinking breather* is a Ricci flow solution on $[t_1, t_2]$ that satisfies $g(t_2) = c\,\phi^* g(t_1)$ for some $c < 1$ and $\phi \in \mathrm{Diff}(M)$.

Gradient shrinking soliton solutions are shrinking breathers.

Again, we are assuming that M is compact.

Proposition 12.18. *A shrinking breather is a gradient shrinking soliton.*

Proof. Put $t_0 = \frac{t_2 - ct_1}{1-c}$. Then if $\tau_1 = t_0 - t_1$ and $\tau_2 = t_0 - t_2$, we have $\tau_2 = c\tau_1$. Hence

$$\mu(g(t_2), \tau_2) = \mu\left(\frac{\tau_2}{\tau_1}\,\phi^* g(t_1), \tau_2\right) = \mu\left(\phi^* g(t_1), \tau_1\right) = \mu\left(g(t_1), \tau_1\right). \tag{12.19}$$

It follows that the solution is a gradient shrinking soliton. □

13. I.4. The no local collapsing theorem I

In this section we prove the no local collapsing theorem.

Definition 13.1. A smooth Ricci flow solution $g(\cdot)$ on a time interval $[0, T)$ is said to be *locally collapsing* at T if there is a sequence of times $t_k \to T$ and a sequence of metric balls $B_k = B(p_k, r_k)$ at times t_k such that r_k^2/t_k is bounded, $|\operatorname{Rm}|(g(t_k)) \le r_k^{-2}$ in B_k and $\lim_{k\to\infty} r_k^{-n}\,\mathrm{vol}(B_k) = 0$.

Remark 13.2. In the definition of noncollapsing, T could be infinite. This is why it is written that r_k^2/t_k stays bounded, while if $T < \infty$ then this is obviously the same as saying that r_k stays bounded.

Theorem 13.3. *(cf. Theorem I.4.1) If M is closed and $T < \infty$ then $g(\cdot)$ is not locally collapsing at T.*

Proof. We first sketch the idea of the proof. In Section 11 we showed that in the case of flat $\mathbb{R}^n$, taking $e^{-f}(x) = e^{-\frac{|x|^2}{4\tau}}$, we get $\mathcal{W}(g, f, \tau) = 0$. So putting $\tau = r_k^2$ and $e^{-f_k}(x) = e^{-\frac{|x|^2}{4r_k^2}}$, we have $\mathcal{W}(g, f_k, r_k^2) = 0$. In the collapsing case, the idea is to use a test function f_k so that

$$e^{-f_k}(x) \sim e^{-c_k}\, e^{-\frac{\mathrm{dist}_{t_k}(x,p_k)^2}{4r_k^2}}, \tag{13.4}$$

where c_k is determined by the normalization condition

$$\int_M (4\pi r_k^2)^{-n/2}\, e^{-f_k}\, dV = 1. \tag{13.5}$$

The main difference between computing (13.5) in M and in $\mathbb{R}^n$ comes from the difference in volumes, which means that $e^{-c_k} \sim \frac{1}{r_k^{-n}\,vol(B_k)}$. In particular, as $k \to \infty$, we have $c_k \to -\infty$.

Now that f_k is normalized correctly, the main difference between computing $\mathcal{W}(g(t_k), f_k, r_k^2)$ in M, and the analogous computation for the Gaussian in $\mathbb{R}^n$, comes from the f term in the integrand of $\mathcal{W}$. Since $f_k \sim c_k$, this will drive $\mathcal{W}(g(t_k), f_k, r_k^2)$ to $-\infty$ as $k \to \infty$, so $\mu(g(t_k), r_k^2) \to -\infty$; by the monotonicity of $\mu(g(t), t_0 - t)$ it follows that $\mu(g(0), t_k + r_k^2) \to -\infty$ as $k \to \infty$. This contradicts the fact that $\mu(g(0), \tau)$ is a continuous function of τ.

To write this out precisely, let us put $\Phi = e^{-f/2}$, so that

$$\mathcal{W}(g, f, \tau) = (4\pi\tau)^{-n/2} \int_M \left[4\tau\, |\nabla \Phi|^2 + (\tau R - 2 \ln \Phi - n)\, \Phi^2\right] dV. \tag{13.6}$$

For the argument, it is enough to obtain small values of $\mathcal{W}$ for positive Φ. Since $\lim_{s\to 0}(-2\ln s)s^2 = 0$, by an approximation it is enough to obtain small values of $\mathcal{W}$ for nonnegative Φ, where the integrand is declared to be $4\tau\, |\nabla\Phi|^2$ at points where Φ vanishes. Take

$$\Phi_k(x) = e^{-c_k/2}\, \phi(\mathrm{dist}_{t_k}(x, p_k)/r_k), \tag{13.7}$$

where $\phi : [0, \infty) \to [0, 1]$ is a monotonically nonincreasing function such that $\phi(s) = 1$ if $s \in [0, 1/2]$, $\phi(s) = 0$ if $s \geq 1$ and $|\phi'(s)| \leq 10$ for $s \in [1/2, 1]$. The function Φ_k is *a priori* only Lipschitz, but by smoothing it slightly we can use Φ_k in the variational formula to bound $\mathcal{W}$ from above.

The constant c_k is determined by

$$e^{c_k} = \int_M (4\pi r_k^2)^{-n/2}\, \phi^2(\mathrm{dist}_{t_k}(x, p_k)/r_k)\, dV \leq (4\pi r_k^2)^{-n/2}\, \mathrm{vol}(B_k). \tag{13.8}$$

Thus $c_k \to -\infty$. Next,

$$\mathcal{W}(g(t_k), f_k, r_k^2) = (4\pi r_k^2)^{-n/2} \int_M \left[4 r_k^2\, |\nabla \Phi_k|^2 + (r_k^2 R - 2 \ln \Phi_k - n)\, \Phi_k^2\right] dV. \tag{13.9}$$

Let $A_k(s)$ be the mass of the distance sphere $S(p_k, r_k s)$ around p_k. Put

$$\overline{R}_k(s) = r_k^2\, A_k(s)^{-1} \int_{S(p_k, r_k s)} R\, d\,\mathrm{area}\,. \tag{13.10}$$

We can compute the integral in (13.9) radially to get

$$\mathcal{W}(g(t_k), f_k, r_k^2) = \frac{\int_0^1 \left[4(\phi'(s))^2 + (\overline{R}_k(s) + c_k - 2\, \ln \phi(s) - n)\, \phi^2(s)\right] A_k(s)\, ds}{\int_0^1 \phi^2(s)\, A_k(s)\, ds}. \tag{13.11}$$

The expression $4(\phi'(s))^2 - 2\, \ln \phi(s)\, \phi^2(s)$ vanishes if $s \notin [1/2, 1]$, and is bounded above by $400 + e^{-1}$ if $s \in [1/2, 1]$. Then the lower bound on the Ricci curvature and the Bishop-Gromov inequality give

$$\begin{aligned} \frac{\int_0^1 [4(\phi'(s))^2 - 2\, \ln \phi(s)\, \phi^2(s)]\, A_k(s)\, ds}{\int_0^1 \phi^2(s)\, A_k(s)\, ds} &\leq 401\, \frac{\mathrm{vol}(B(p_k, r_k)) - \mathrm{vol}(B(p_k, r_k/2))}{\mathrm{vol}(B(p_k, r_k/2))} \\ &\leq 401 \left(\frac{\int_0^1 \sinh^{n-1}(s)\, ds}{\int_0^{1/2} \sinh^{n-1}(s)\, ds} - 1 \right). \end{aligned} \tag{13.12}$$

Next, from the upper bound on scalar curvature, $\overline{R}_k(s) \leq n(n-1)$ for $s \in [0, 1]$. Putting this together gives $\mathcal{W}(g(t_k), f_k, r_k^2) \leq \mathrm{const.} + c_k$ and so $\mathcal{W}(g(t_k), f_k, r_k^2) \to -\infty$ as $k \to \infty$.

Thus $\mu(g(t_k), r_k^2) \to -\infty$. For any $t_0 > t$, $\mu(g(t), t_0 - t)$ is nondecreasing in t. Hence $\mu(g(0), t_k + r_k^2) \le \mu(g(t_k), r_k^2)$, so $\mu(g(0), t_k + r_k^2) \to -\infty$. Since T is finite, t_k and r_k^2 are uniformly bounded, and t_k uniformly positive, which contradicts the fact that $\mu(g(0), \tau)$ is a continuous function of τ. □

Remark 13.13. In the preceding argument we only used the upper bound on scalar curvature and the lower bound on Ricci curvature, i.e. in the definition of local collapsing one could have assumed that $R(g_{ij}(t_k)) \le n(n-1)\, r_k^{-2}$ in B_k and $\mathrm{Ric}(g_{ij}(t_k)) \ge -(n-1)\, r_k^{-2}$ in B_k. In fact, one can also remove the lower bound on Ricci curvature (observation of Perelman, communicated by Gang Tian). The necessary ingredients of the preceding argument were that
1. $r_k^{-n}\ \mathrm{vol}(B(p_k, r_k)) \to 0$,
2. $r_k^2\, R$ is uniformly bounded above on $B(p_k, r_k)$ and
3. $\frac{\mathrm{vol}(B(p_k, r_k))}{\mathrm{vol}(B(p_k, r_k/2))}$ is uniformly bounded above.

Suppose only that $r_k^{-n}\ \mathrm{vol}(B(p_k, r_k)) \to 0$ and for all k, $r_k^2\, R \le n(n-1)$ on $B(p_k, r_k)$. If $\frac{\mathrm{vol}(B(p_k, r_k))}{\mathrm{vol}(B(p_k, r_k/2))} < 3^n$ for all k then we are done. If not, suppose that for a given k, $\frac{\mathrm{vol}(B(p_k, r_k))}{\mathrm{vol}(B(p_k, r_k/2))} \ge 3^n$. Putting $r_k' = r_k/2$, we have that $(r_k')^{-n}\ \mathrm{vol}(B(p_k, r_k')) \le r_k^{-n}\ \mathrm{vol}(B(p_k, r_k))$ and $(r_k')^2\, R \le n(n-1)$ on $B(p_k, r_k')$. We replace r_k by r_k'. If now $\frac{\mathrm{vol}(B(p_k, r_k))}{\mathrm{vol}(B(p_k, r_k/2))} < 3^n$ then we stop. If not then we repeat the process and replace r_k by $r_k/2$. Eventually we will achieve that $\frac{\mathrm{vol}(B(p_k, r_k))}{\mathrm{vol}(B(p_k, r_k/2))} < 3^n$. Then we can apply the preceding argument to this new sequence of pairs $\{(p_k, r_k)\}_{k=1}^{\infty}$.

Definition 13.14. (cf. Definition I.4.2) We say that a metric g is *κ-noncollapsed on the scale ρ* if every metric ball B of radius $r < \rho$, which satisfies $|\mathrm{Rm}(x)| \le r^{-2}$ for every $x \in B$, has volume at least κr^n.

Remark 13.15. We caution the reader that this definition differs slightly from the definition of noncollapsing that is used from section I.7 onwards.

Note that except for the overall scale ρ, the κ-noncollapsed condition is scale-invariant. From the proof of Theorem 13.3 we extract the following statement. Given a Ricci flow defined on an interval $[0, T)$, with $T < \infty$, and a scale ρ, there is some number $\kappa = \kappa(g(0), T, \rho)$ so that the solution is κ-noncollapsed on the scale ρ for all $t \in [0, T)$. We note that the estimate on κ deteriorates as $T \to \infty$, as there are Ricci flow solutions that collapse at long time.

14. I.5. The $\mathcal{W}$-functional as a time derivative

We will only discuss one formula from I.5, showing that along a Ricci flow, $\mathcal{W}$ is itself the time-derivative of an integral expression.

Again, we put $\tau = -t$. Consider the evolution equations (12.9), with $(4\pi\tau)^{-n/2}\, e^{-f}\, dV = dm$. Then

$$\begin{aligned} \frac{d}{d\tau}\left(\tau \int_M \left(f - \frac{n}{2}\right) dm\right) &= \int_M \left(f - \frac{n}{2}\right) dm + \tau \int_M \left(\triangle f + R - \frac{n}{2\tau}\right) dm \\ &= \int_M \left(f - \frac{n}{2}\right) dm + \tau \int_M \left(|\nabla f|^2 + R - \frac{n}{2\tau}\right) dm \\ &= \mathcal{W}(g(t), f(t), \tau). \end{aligned} \tag{14.1}$$

With respect to the evolution equations (12.12) obtained by performing diffeomorphisms, we get

$$\frac{d}{d\tau}\left(\tau \int_M \left(f - \frac{n}{2}\right) (4\pi\tau)^{-n/2}\, e^{-f}\, dV\right) = \mathcal{W}(g(t), f(t), \tau). \tag{14.2}$$

Similarly, with respect to (5.21),

$$\frac{d}{dt}\left(-\int_M f\, e^{-f}\, dV\right) = \mathcal{F}(g(t), f(t)). \tag{14.3}$$

15. I.7. Overview of reduced length and reduced volume

We first give a brief summary of I.7. In I.7, the variable $\tau = t_0 - t$ is used and so the corresponding Ricci flow equation is $(g_{ij})_\tau = 2R_{ij}$. The goal is to prove a no local collapsing theorem by means of the $\mathcal{L}$-lengths of curves $\gamma : [\tau_1, \tau_2] \to M$, defined by

$$\mathcal{L}(\gamma) = \int_{\tau_1}^{\tau_2} \sqrt{\tau}\, \left(R(\gamma(\tau)) + \left|\dot{\gamma}(\tau)\right|^2\right) d\tau, \tag{15.1}$$

where the scalar curvature $R(\gamma(\tau))$ and the norm $|\dot{\gamma}(\tau)|$ are evaluated using the metric at time $t_0 - \tau$. Here $\tau_1 \geq 0$. With $X = \frac{d\gamma}{d\tau}$, the corresponding $\mathcal{L}$-geodesic equation is

$$\nabla_X X - \frac{1}{2}\nabla R + \frac{1}{2\tau} X + 2\,\mathrm{Ric}(X, \cdot) = 0, \tag{15.2}$$

where again the connection and curvature are taken at the corresponding time, and the 1-form $\mathrm{Ric}(X, \cdot)$ has been identified with the corresponding dual vector field.

Fix $p \in M$. Taking $\tau_1 = 0$ and $\gamma(0) = p$, the vector $v = \lim_{\tau \to 0} \sqrt{\tau}\, X(\tau)$ is well-defined in T_pM and is called the initial vector of the geodesic. The $\mathcal{L}$-exponential map $\mathcal{L}exp_\tau : T_pM \to M$ sends v to $\gamma(\tau)$.

The function $L(q, \overline{\tau})$ is the infimal $\mathcal{L}$-length of curves γ with $\gamma(0) = p$ and $\gamma(\overline{\tau}) = q$. Defining the reduced length by

$$l(q, \tau) = \frac{L(q, \tau)}{2\sqrt{\tau}} \tag{15.3}$$

and the reduced volume by

$$\tilde{V}(\tau) = \int_M \tau^{-\frac{n}{2}}\, e^{-l(q,\tau)}\, dq, \tag{15.4}$$

the goal is to show that $\tilde{V}(\tau)$ is nonincreasing in τ, i.e. nondecreasing in t. To do this, one uses the $\mathcal{L}$-exponential map to write $\tilde{V}(\tau)$ as an integral over T_pM:

$$\tilde{V}(\tau) \;=\; \int_{T_pM} \tau^{-\frac{n}{2}}\, e^{-l(\mathcal{L}exp_\tau(v),\tau)}\, \mathcal{J}(v,\tau)\, \chi_\tau(v)\, dv, \tag{15.5}$$

where $\mathcal{J}(v,\tau) \;=\; \det\, d\,(\mathcal{L}exp_\tau)_v$ is the Jacobian factor in the change of variable and χ_τ is a cutoff function related to the $\mathcal{L}$-cut locus of p. To show that $\tilde{V}(\tau)$ is nonincreasing in τ it suffices to show that $\tau^{-\frac{n}{2}}\, e^{-l(\mathcal{L}exp_\tau(v),\tau)}\, \mathcal{J}(v,\tau)$ is nonincreasing in τ or, equivalently, that $-\frac{n}{2}\,\ln(\tau) \;-\; l(\mathcal{L}exp_\tau(v),\tau) \;+\; \ln\,\mathcal{J}(v,\tau)$ is nonincreasing in τ. Hence it is necessary to compute $\frac{dl(\mathcal{L}exp_\tau(v),\tau)}{d\tau}$ and $\frac{d\mathcal{J}(v,\tau)}{d\tau}$. The computation of the latter will involve the $\mathcal{L}$-Jacobi fields.

The fact that $\tilde{V}(\tau)$ is nonincreasing in τ is then used to show that the Ricci flow solution cannot be collapsed near p.

16. Basic example for I.7

In this section we say what the various expressions of I.7 become in the model case of a flat Euclidean Ricci solution.

If M is flat $\mathbb{R}^n$ and $p \;=\; \vec{0}$ then the unique $\mathcal{L}$-geodesic γ with $\gamma(0) \;=\; \vec{0}$ and $\gamma(\overline{\tau}) \;=\; \vec{q}$ is

$$\gamma(\tau) \;=\; \left(\frac{\tau}{\overline{\tau}}\right)^{\frac{1}{2}}\, \vec{q} \;=\; 2\,\tau^{\frac{1}{2}}\,\vec{v}. \tag{16.1}$$

The function L is given by

$$L(q,\overline{\tau}) \;=\; \frac{1}{2}\,\overline{\tau}^{\,-\frac{1}{2}}\,|q|^2 \tag{16.2}$$

and the reduced length (15.3) is given by

$$l(q,\overline{\tau}) \;=\; \frac{|q|^2}{4\overline{\tau}}. \tag{16.3}$$

The function $\overline{L}(q,\overline{\tau}) \;=\; 2\,\overline{\tau}^{\frac{1}{2}}\, L(q,\overline{\tau})$ is

$$\overline{L}(q,\overline{\tau}) \;=\; |q|^2. \tag{16.4}$$

Then

$$\tilde{V}(\tau) \;=\; \int_{\mathbb{R}^n} \tau^{-\frac{n}{2}}\, e^{-\frac{|q|^2}{4\tau}}\, d^nq \;=\; (4\pi)^{\frac{n}{2}} \tag{16.5}$$

is constant in τ.

17. Remarks about $\mathcal{L}$-Geodesics and $\mathcal{L}\exp$

In this section we discuss the variational equation corresponding to (15.1).

To derive the $\mathcal{L}$-geodesic equation, as in Riemannian geometry we consider a 1-parameter family of curves $\gamma_s : [\tau_1,\tau_2] \to M$, parametrized by $s \in (-\epsilon,\epsilon)$. Equivalently, we have a map $\tilde{\gamma}(s,\tau)$ with $s \in (-\epsilon,\epsilon)$ and $t \in [\tau_1,\tau_2]$. Putting $X \;=\; \frac{\partial\tilde{\gamma}}{\partial\tau}$ and $Y \;=\; \frac{\partial\tilde{\gamma}}{\partial s}$, we have $[X,Y] \;=\; 0$.

Then $\nabla_X Y = \nabla_Y X$. Restricting to the curve $\gamma(\tau) = \tilde{\gamma}(0,\tau)$ and writing δ_Y as shorthand for $\frac{d}{ds}\big|_{s=0}$, we have $(\delta_Y \gamma)(\tau) = Y(\tau)$ and $(\delta_Y X)(\tau) = (\nabla_X Y)(\tau)$. Then

$$\delta_Y \mathcal{L} = \int_{\tau_1}^{\tau_2} \sqrt{\tau}\, \left(\langle Y, \nabla R \rangle + 2 \langle \nabla_X Y, X \rangle \right) d\tau. \tag{17.1}$$

Using the fact that $\frac{dg_{ij}}{d\tau} = 2R_{ij}$, we have

$$\frac{d\langle Y, X\rangle}{d\tau} = \langle \nabla_X Y, X \rangle + \langle Y, \nabla_X X \rangle + 2 \operatorname{Ric}(Y, X). \tag{17.2}$$

Then

$$\begin{aligned} &\int_{\tau_1}^{\tau_2} \sqrt{\tau}\, \left(\langle Y, \nabla R \rangle + 2 \langle \nabla_X Y, X \rangle \right) d\tau = \\ &\int_{\tau_1}^{\tau_2} \sqrt{\tau} \left(\langle Y, \nabla R \rangle + 2 \frac{d}{d\tau} \langle Y, X \rangle - 2 \langle Y, \nabla_X X \rangle - 4 \operatorname{Ric}(Y, X) \right) d\tau = \\ &2 \sqrt{\tau}\, \langle X, Y \rangle \Big|_{\tau_1}^{\tau_2} + \int_{\tau_1}^{\tau_2} \sqrt{\tau} \left\langle Y, \nabla R - 2 \nabla_X X - 4 \operatorname{Ric}(X, \cdot) - \frac{1}{\tau} X \right\rangle d\tau. \end{aligned} \tag{17.3}$$

Hence the $\mathcal{L}$-geodesic equation is

$$\nabla_X X - \frac{1}{2} \nabla R + \frac{1}{2\tau} X + 2 \operatorname{Ric}(X, \cdot) = 0. \tag{17.4}$$

We now discuss some technical issues about $\mathcal{L}$-geodesics and the $\mathcal{L}$-exponential map. We are assuming that $(M, g(\cdot))$ is a Ricci flow, where the curvature operator of M is uniformly bounded on a τ-interval $[\tau_1, \tau_2]$, and each τ-slice $(M, g(\tau))$ is complete for $\tau \in [\tau_1, \tau_2]$. By Appendix D, for every $\tau' < \tau_2$ there is a constant $D < \infty$ such that

$$|\nabla R(x, \tau)| < \frac{D}{\sqrt{\tau_2 - \tau}} \tag{17.5}$$

for all $x \in M$, $\tau \in [\tau_1, \tau_2)$.

Making the change of variable $s = \sqrt{\tau}$ in the formula for $\mathcal{L}$-length, we get

$$\mathcal{L}(\gamma) = 2 \int_{s_1}^{s_2} \left(\frac{1}{4} \left| \frac{d\gamma}{ds} \right|^2 + s^2 R(\gamma(s)) \right) ds. \tag{17.6}$$

The Euler-Lagrange equation becomes

$$\nabla_{\hat{X}} \hat{X} - 2s^2 \nabla R + 4s \operatorname{Ric}(\hat{X}, \cdot) = 0, \tag{17.7}$$

where $\hat{X} = \frac{d\gamma}{ds} = 2sX$. Putting $s_1 = \sqrt{\tau_1}$, it follows from standard existence theory for ODE's that for each $p \in M$ and $v \in T_p M$, there is a unique solution $\gamma(s)$ to (17.7), defined on an interval $[s_1, s_1 + \epsilon)$, with $\gamma(s_1) = p$ and

$$\frac{1}{2} \gamma'(s_1) = \lim_{\tau \to \tau_1} \sqrt{\tau} \frac{d\gamma}{d\tau} = v. \tag{17.8}$$

If $\gamma(s)$ is defined for $s \in [s_1, s']$ then

$$\begin{aligned} \frac{d}{ds} |\hat{X}|^2 = \frac{d}{ds} \langle \hat{X}, \hat{X} \rangle &= 4s \operatorname{Ric}(\hat{X}, \hat{X}) + 2 \langle \nabla_{\hat{X}} \hat{X}, \hat{X} \rangle \\ &= -4s \operatorname{Ric}(\hat{X}, \hat{X}) + 4s^2 \langle \nabla R, \hat{X} \rangle \end{aligned} \tag{17.9}$$

and so if $\hat{X}(s) \neq 0$ then

$$\frac{d}{ds}|\hat{X}| = \frac{1}{2|\hat{X}|}\frac{d}{ds}|\hat{X}|^2 \;=\; -2s|\hat{X}|\operatorname{Ric}\left(\frac{\hat{X}}{|\hat{X}|}, \frac{\hat{X}}{|\hat{X}|}\right) + 2s^2\left\langle \nabla R, \frac{\hat{X}}{|\hat{X}|}\right\rangle. \tag{17.10}$$

By (17.5),

$$\frac{d}{ds}|\hat{X}| \;\leq\; C_1|\hat{X}| + \frac{C_2}{\sqrt{s_2 - s}} \tag{17.11}$$

for appropriate constants C_1 and C_2, where $s_2 = \sqrt{\tau_2}$. Since the metrics $g(\tau)$ are uniformly comparable for $\tau \in [\tau_1, \tau_2]$, we conclude (by a continuity argument in s) that the $\mathcal{L}$-geodesic γ_v with $\frac{1}{2}\gamma_v'(s_1) = v$ is defined on the whole interval $[s_1, s_2]$. In particular, in terms of the original variable τ, for each $\tau \in [\tau_1, \tau_2]$ and each $p \in M$, we get a globally defined and smooth $\mathcal{L}$-exponential map $\mathcal{L}\exp_\tau : T_pM \to M$ which takes each $v \in T_pM$ to $\gamma(\tau)$, where $v = \lim_{\tau' \to \tau_1} \sqrt{\tau'}\frac{d\gamma}{d\tau'}$. Note that unlike in the case of Riemannian geometry, $\mathcal{L}\exp_\tau(0)$ may not be p, because of the ∇R term in (17.4).

We now fix $p \in M$, take $\tau_1 = 0$, and let $L(q, \bar{\tau})$ be the minimizer function as in Section 15. We can imitate the traditional Riemannian geometry proof that geodesics minimize for a short time. Using the change of variable $s = \sqrt{\tau}$ and the implicit function theorem, there is an $r = r(p) > 0$ (which varies continuously with p) such that for every $q \in M$ with $d(q, p) \leq 10r$ at $\tau = 0$, and every $0 < \bar{\tau} \leq r^2$, there is a unique $\mathcal{L}$-geodesic $\gamma_{(q,\bar{\tau})} : [0, \bar{\tau}] \to M$, starting at p and ending at q, which remains within the ball $B(p, 100r)$ (in the $\tau = 0$ slice $(M, g(0))$), and $\gamma_{(q,\bar{\tau})}$ varies smoothly with $(q, \bar{\tau})$. Thus, the $\mathcal{L}$-length of $\gamma_{(q,\bar{\tau})}$ varies smoothly with $(q, \bar{\tau})$, and defines a function $\hat{L}(q, \bar{\tau})$ near $(p, 0)$. We claim that $\hat{L} = L$ near $(p, 0)$. Suppose that $q \in B(p, r)$ and let $\alpha : [0, \bar{\tau}] \to M$ be a smooth curve whose $\mathcal{L}$-length is close to $L(q, \bar{\tau})$. If r is small, relative to the assumed curvature bound, then α must stay within $B(p, 10r)$. Equations (18.2) and (18.6) below imply that

$$\frac{d}{d\tau}\hat{L}(\alpha(\tau), \tau) \;=\; \langle 2\sqrt{\tau}X, \frac{d\alpha}{d\tau}\rangle + \sqrt{\tau}(R - |X|^2) \;\leq\; \sqrt{\tau}\left(R + \left|\frac{d\alpha}{d\tau}\right|^2\right) \;=\; \frac{d}{d\tau}\left(\mathcal{L}\mathrm{length}(\alpha\big|_{[0,\tau]})\right). \tag{17.12}$$

Thus $\gamma_{(q,\bar{\tau})}$ minimizes when $(q, \bar{\tau})$ is close to $(p, 0)$.

We can now deduce that for all $(q, \bar{\tau})$, there is an $\mathcal{L}$-geodesic $\gamma : [0, \bar{\tau}] \to M$ which has infimal $\mathcal{L}$-length among all piecewise smooth curves starting at p and ending at q (with domain $[0, \bar{\tau}]$). This can be done by imitating the usual broken geodesic argument, using the fact that for x, y in a given small ball of M and for sufficiently small time intervals $[\tau', \tau' + \epsilon] \subset [0, \overline{\tau}]$, there is a unique minimizer γ for $\int_{\tau'}^{\tau'+\epsilon} \sqrt{\tau}\left(R(\gamma(\tau)) \;+\; |\dot{\gamma}(\tau)|^2\right)\, d\tau$ with $\gamma(\tau') = x$ and $\gamma(\tau' + \epsilon) = y$. Alternatively, using the change of variable $s = \sqrt{\tau}$, one can take a minimizer of $\mathcal{L}$ among $H^{1,2}$-regular curves.

Another technical issue is the justification of the change of variables from M to T_pM in the proof of monotonicity of reduced volume. Fix $p \in M$ and $\tau > 0$, and let $\mathcal{L}\exp_\tau : T_pM \to M$ be the map which takes $v \in T_pM$ to $\gamma_v(\tau)$, where $\gamma_v : [0, \tau] \to M$ is the unique $\mathcal{L}$-geodesic with $\sqrt{\tau'}\frac{d\gamma_v}{d\tau'} \to v$ as $\tau' \to 0$. Let $\mathcal{B} \subset M$ be the set of points which are either endpoints of more than one minimizing $\mathcal{L}$-geodesic, or which are the endpoint of a minimizing geodesic $\gamma_v : [0, \tau] \to M$ where $v \in T_pM$ is a critical point of $\mathcal{L}\exp_\tau$. We will call $\mathcal{B}$ the time-τ $\mathcal{L}$-cut

locus of p. It is a closed subset of M. Let $\mathcal{G} \subset M$ be the complement of $\mathcal{B}$ and let $\Omega_\tau \subset T_pM$ be the corresponding set of initial conditions for minimizing $\mathcal{L}$-geodesics. Then Ω_τ is an open set, and $\mathcal{L}\exp$ maps it diffeomorphically onto $\mathcal{G}$. We claim that $\mathcal{B}$ has measure zero. By Sard's theorem, to prove this it suffices to prove that the set $\mathcal{B}'$ of points $q \in \mathcal{B}$ which are regular values of $\mathcal{L}\exp_\tau$, has measure zero. Pick $q \in \mathcal{B}'$, and distinct points $v_1, v_2 \in T_pM$ such that $\gamma_{v_i} : [0, \tau] \to M$ are both minimizing geodesics ending at q. Then $\mathcal{L}\exp_\tau$ is a local diffeomorphism near each v_i. The first variation formula and the implicit function theorem then show that there are neighborhoods U_i of v_i, and a smooth hypersurface H passing through q, such that if we have points $w_i \in U_i$ with

$$q' = \mathcal{L}\exp_\tau(w_1) = \mathcal{L}\exp_\tau(w_2) \quad \text{and} \quad \mathcal{L}\,\mathrm{length}(\gamma_{w_1}) = \mathcal{L}\,\mathrm{length}(\gamma_{w_2}), \tag{17.13}$$

then q' lies on H. Thus $\mathcal{B}'$ is contained in a countable union of hypersurfaces, and hence has measure zero.

Therefore one may compute the integral of any integrable function on M by pulling it back to $\Omega_\tau \subset T_pM$ and using the change of variables formula. Note that if $\tau \leq \tau'$ then $\Omega_{\tau'} \subset \Omega_\tau$.

18. I.(7.3)-(7.6). First derivatives of L

In this section we do some preliminary calculations leading up to the computation of the second variation of L.

A remark about the notation : L is a function of a point q and a time $\overline{\tau}$. The notation $L_{\overline{\tau}}$ refers to the partial derivative with respect to $\overline{\tau}$, i.e. differentiation while keeping q fixed. The notation $\frac{d}{d\tau}$ refers to differentiation along an $\mathcal{L}$-geodesic, i.e. simultaneously varying both the point and the time.

If q is not in the time-$\overline{\tau}$ $\mathcal{L}$-cut locus of p, let $\gamma : [0, \overline{\tau}] \to M$ be the unique minimizing $\mathcal{L}$-geodesic from p to q, with length $L(q, \overline{\tau})$. If $c : (-\epsilon, \epsilon) \to M$ is a short curve with $c(0) = q$, consider the 1-parameter family of minimizing $\mathcal{L}$-geodesics $\tilde{\gamma}(s, \tau)$ with $\tilde{\gamma}(s, 0) = p$ and $\tilde{\gamma}(s, \overline{\tau}) = c(s)$. Putting $Y(\tau) = \frac{\partial \tilde{\gamma}(s,\tau)}{\partial s}\Big|_{s=0}$, equation (17.3) gives

$$\langle \nabla L, c'(0) \rangle = \frac{dL(c(s), \overline{\tau})}{ds}\Big|_{s=0} = 2\sqrt{\overline{\tau}}\, \langle X(\overline{\tau}), Y(\overline{\tau}) \rangle. \tag{18.1}$$

Hence

$$(\nabla L)(q, \overline{\tau}) = 2\sqrt{\overline{\tau}}\, X(\overline{\tau}) \tag{18.2}$$

and

$$|\nabla L|^2(q, \overline{\tau}) = 4\,\overline{\tau}\, |X(\overline{\tau})|^2 = -\,4\,\overline{\tau}\, R(q) + 4\,\overline{\tau}\, \left(R(q) + |X(\overline{\tau})|^2 \right). \tag{18.3}$$

If we simply extend the $\mathcal{L}$-geodesic γ in $\overline{\tau}$, we obtain

$$\frac{dL(\gamma(\overline{\tau}), \overline{\tau})}{d\overline{\tau}} = \sqrt{\overline{\tau}}\, \left(R(\gamma(\tau)) + |X(\overline{\tau})|^2 \right). \tag{18.4}$$

As

$$\frac{dL(\gamma(\overline{\tau}), \overline{\tau})}{d\overline{\tau}} = L_{\overline{\tau}}(q, \overline{\tau}) + \langle (\nabla L)(q, \overline{\tau}), X(\overline{\tau}) \rangle, \tag{18.5}$$

equations (18.2) and (18.4) give

$$L_{\overline{\tau}}(q,\overline{\tau}) \;=\; \sqrt{\overline{\tau}}\,\left(R(q) \,+\, |X(\overline{\tau})|^2\right) \,-\, \langle (\nabla L)(q,\overline{\tau}), X(\overline{\tau})\rangle \tag{18.6}$$
$$= 2\sqrt{\overline{\tau}} R(q) \,-\, \sqrt{\overline{\tau}}\,\left(R(q) \,+\, |X(\overline{\tau})|^2\right).$$

When computing $\frac{dl(\gamma(\tau),\tau)}{d\tau}$, it will be useful to have a formula for $R(\gamma(\tau)) \,+\, \left|X(\tau)\right|^2$. As in I.(7.3),

$$\frac{d}{d\tau}\left(R(\gamma(\tau)) \,+\, \left|X(\tau)\right|^2\right) \;=\; R_\tau \,+\, \langle \nabla R, X\rangle \,+\, 2\langle \nabla_X X, X\rangle \,+\, 2\,\mathrm{Ric}(X,X). \tag{18.7}$$

Using the $\mathcal{L}$-geodesic equation (17.4) gives

(18.8)
$$\frac{d}{d\tau}\left(R(\gamma(\tau)) \,+\, \left|X(\tau)\right|^2\right) \;=\; R_\tau \,+\, \frac{1}{\tau}\,R \,+\, 2\langle \nabla R, X\rangle \,-\, 2\,\mathrm{Ric}(X,X) \,-\, \frac{1}{\tau}\,(R+|X|^2)$$
$$= -\,H(X) \,-\, \frac{1}{\tau}\,(R+|X|^2),$$

where

$$H(X) \;=\; -\,R_\tau \,-\, \frac{1}{\tau}\,R \,-\, 2\langle \nabla R, X\rangle \,+\, 2\,\mathrm{Ric}(X,X) \tag{18.9}$$

is the expression of (F.9) after the change $\tau \,=\, -t$ and $X \to -X$. Multiplying (18.8) by $\tau^{\frac32}$ and integrating gives

$$\int_0^{\overline{\tau}} \tau^{\frac32}\,\frac{d}{d\tau}\left(R(\gamma(\tau)) \,+\, \left|X(\tau)\right|^2\right)\,d\tau \;=\; -\,K \,-\, L(q,\overline{\tau}), \tag{18.10}$$

where

$$K \;=\; \int_0^{\overline{\tau}} \tau^{\frac32}\,H(X(\tau))\,d\tau. \tag{18.11}$$

Then integrating the left-hand side of (18.10) by parts gives

$$\overline{\tau}^{\frac32}\,\left(R(\gamma(\overline{\tau})) \,+\, \left|X(\overline{\tau})\right|^2\right) \;=\; -\,K \,+\, \frac{1}{2}\,L(q,\overline{\tau}). \tag{18.12}$$

Plugging this back into (18.6) and (18.3) gives

$$L_{\overline{\tau}}(q,\overline{\tau}) \;=\; 2\sqrt{\overline{\tau}}R(q) \,-\, \frac{1}{2\overline{\tau}}\,L(q,\overline{\tau}) \,+\, \frac{1}{\overline{\tau}}\,K \tag{18.13}$$

and

$$|\nabla L|^2(q,\overline{\tau}) \;=\; -\,4\,\overline{\tau}\,R(q) \,+\, \frac{2}{\sqrt{\overline{\tau}}}\,L(q,\overline{\tau}) \,-\, \frac{4}{\sqrt{\overline{\tau}}}\,K. \tag{18.14}$$

19. I.(7.7). Second variation of $\mathcal{L}$

In this section we compute the second variation of $\mathcal{L}$. We use it to compute the Hessian of L on M.

To compute the second variation $\delta_Y^2\mathcal{L}$, we start with the first variation equation

$$\delta_Y \mathcal{L} = \int_0^{\overline{\tau}} \sqrt{\tau}\ (\langle Y, \nabla R\rangle + 2\langle \nabla_Y X, X\rangle)\, d\tau. \tag{19.1}$$

Recalling that $\delta_Y\gamma(\tau) = Y(\tau)$ and $\delta_Y X(\tau) = (\nabla_Y X)(\tau)$, the second variation is

(19.2)
$$\begin{aligned}
\delta_Y^2\mathcal{L} &= \int_0^{\overline{\tau}} \sqrt{\tau}\ (Y\cdot Y\cdot R + 2\langle \nabla_Y\nabla_Y X, X\rangle + 2\langle \nabla_Y X, \nabla_Y X\rangle)\, d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\ \left(Y\cdot Y\cdot R + 2\langle \nabla_Y\nabla_X Y, X\rangle + 2\left|\nabla_X Y\right|^2\right) d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\ \left(Y\cdot Y\cdot R + 2\langle \nabla_X\nabla_Y Y, X\rangle + 2\langle R(Y,X)Y, X\rangle + 2\left|\nabla_X Y\right|^2\right) d\tau,
\end{aligned}$$

where the notation $Z\cdot u$ refers to the directional derivative, i.e. $Z\cdot u = i_Z\, du$. In order to deal with the $\langle \nabla_X\nabla_Y Y, X\rangle$ term, we have to compute $\frac{d}{d\tau}\langle \nabla_Y Y, X\rangle$.

From the general equation for the Levi-Civita connection in terms of the metric [17, (1.29)], if $g(\tau)$ is a 1-parameter family of metrics, with $\dot{g} = \frac{dg}{d\tau}$ and $\dot{\nabla} = \frac{d\nabla}{d\tau}$, then

$$2\langle \dot{\nabla}_X Y, Z\rangle = (\nabla_X \dot{g})(Y,Z) + (\nabla_Y \dot{g})(Z,X) - (\nabla_Z \dot{g})(X,Y). \tag{19.3}$$

In our case $\dot{g} = 2\ \mathrm{Ric}$ and so

$$\begin{aligned}
\frac{d}{d\tau}\langle \nabla_Y Y, X\rangle &= \langle \nabla_X\nabla_Y Y, X\rangle + \langle \nabla_Y Y, \nabla_X X\rangle + 2\,\mathrm{Ric}(\nabla_Y Y, X) + \langle \dot{\nabla}_Y Y, X\rangle \\
&= \langle \nabla_X\nabla_Y Y, X\rangle + \langle \nabla_Y Y, \nabla_X X\rangle + \\
&\quad 2\,\mathrm{Ric}(\nabla_Y Y, X) + 2(\nabla_Y \mathrm{Ric})(Y,X) - (\nabla_X \mathrm{Ric})(Y,Y).
\end{aligned} \tag{19.4}$$

(Although we will not need it, we can write

(19.5)
$$\begin{aligned}
2Y\cdot \mathrm{Ric}(Y,X) - X\cdot \mathrm{Ric}(Y,Y) &= 2(\nabla_Y \mathrm{Ric})(Y,X) + 2\,\mathrm{Ric}(\nabla_Y Y, X) + 2\,\mathrm{Ric}(Y, \nabla_Y X) \\
&\quad - (\nabla_X \mathrm{Ric})(Y,Y) - 2\,\mathrm{Ric}(\nabla_X Y, Y) \\
&= 2\,\mathrm{Ric}(\nabla_Y Y, X) + 2(\nabla_Y \mathrm{Ric})(Y,X) \\
&\quad - (\nabla_X \mathrm{Ric})(Y,Y) - 2\,\mathrm{Ric}([X,Y], Y).
\end{aligned}$$

We are assuming that the variation field Y satisfies $[X,Y] = 0$ (this was used in deriving the $\mathcal{L}$-geodesic equation). Hence one obtains the formula

$$\frac{d}{d\tau}\langle \nabla_Y Y, X\rangle = \langle \nabla_X\nabla_Y Y, X\rangle + \langle \nabla_Y Y, \nabla_X X\rangle + 2Y\cdot \mathrm{Ric}(Y,X) - X\cdot \mathrm{Ric}(Y,Y) \tag{19.6}$$

of I.7.)

Next, using (19.4),

(19.7)

$$
\begin{aligned}
2\sqrt{\overline{\tau}}\langle \nabla_Y Y, X\rangle &= 2\int_0^{\overline{\tau}} \frac{d}{d\tau}\left(\sqrt{\tau}\langle \nabla_Y Y, X\rangle\right)\, d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\left[\frac{1}{\tau}\,\langle \nabla_Y Y, X\rangle \,+\, 2\,\frac{d}{d\tau}\,\langle \nabla_Y Y, X\rangle\right]\, d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\left[\frac{1}{\tau}\,\langle \nabla_Y Y, X\rangle \,+\, 2\langle \nabla_X \nabla_Y Y, X\rangle \,+\, 2\langle \nabla_Y Y, \nabla_X X\rangle + \right.\\
&\qquad\qquad \left. 4\operatorname{Ric}(\nabla_Y Y, X) \,+\, 4(\nabla_Y \operatorname{Ric})(Y,X) \,-\, 2(\nabla_X \operatorname{Ric})(Y,Y)\right]\, d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\,[2\langle \nabla_X \nabla_Y Y, X\rangle \,+\, (\nabla_Y Y)R + \\
&\qquad\qquad 4(\nabla_Y \operatorname{Ric})(Y,X) \,-\, 2(\nabla_X \operatorname{Ric})(Y,Y)]\, d\tau \\
&\quad - \int_0^{\overline{\tau}} \sqrt{\tau}\left\langle \nabla_Y Y, \nabla R \,-\, 2\nabla_X X \,-\, 4\,\operatorname{Ric}(X,\cdot) \,-\, \frac{1}{\tau}\,X\right\rangle\, d\tau.
\end{aligned}
$$

(Of course the last term vanishes if γ is an $\mathcal{L}$-geodesic, but we do not need to assume this here.)

The quadratic form Q representing the Hessian of $\mathcal{L}$ on the path space is given by

$$
\begin{aligned}
Q(Y,Y) &= \delta_Y^2\mathcal{L} \,-\, \delta_{\nabla_Y Y}\mathcal{L} \\
&= \delta_Y^2\mathcal{L} \,-\, 2\sqrt{\overline{\tau}}\langle \nabla_Y Y, X\rangle - \\
&\qquad \int_0^{\overline{\tau}} \sqrt{\tau}\left\langle \nabla_Y Y, \nabla R \,-\, 2\nabla_X X \,-\, 4\,\operatorname{Ric}(X,\cdot) \,-\, \frac{1}{\tau}\,X\right\rangle.
\end{aligned}
\tag{19.8}
$$

It follows that

$$
\begin{aligned}
Q(Y,Y) &= \int_0^{\overline{\tau}} \sqrt{\tau}\,[Y\cdot Y\cdot R \,-\, (\nabla_Y Y)R \,+\, 2\langle R(Y,X)Y, X\rangle + \\
&\qquad 2\,|\nabla_X Y|^2 \,-\, 4(\nabla_Y \operatorname{Ric})(Y,X) \,+\, 2(\nabla_X \operatorname{Ric})(Y,Y)\big]\, d\tau \\
&= \int_0^{\overline{\tau}} \sqrt{\tau}\left[\operatorname{Hess}_R(Y,Y) \,+\, 2\langle R(Y,X)Y, X\rangle \,+\, 2\,|\nabla_X Y|^2\right. \\
&\qquad\qquad \left. -\, 4(\nabla_Y \operatorname{Ric})(Y,X) \,+\, 2(\nabla_X \operatorname{Ric})(Y,Y)\right]\, d\tau.
\end{aligned}
\tag{19.9}
$$

There is an associated second-order differential operator T on vector fields Y given by saying that $2\sqrt{\tau}\langle Y, TY\rangle$ equals the integrand of (19.9) minus $2\frac{d}{d\tau}\left(\sqrt{\tau}\,\langle \nabla_X Y, Y\rangle\right)$. Explicitly,

$$
TY \,=\, -\nabla_X\nabla_X Y - \frac{1}{2\tau}\nabla_X Y + \frac{1}{2}\,\operatorname{Hess}_R(Y,\cdot) - 2\,(\nabla_Y \operatorname{Ric})(X,\cdot) - 2\,\operatorname{Ric}(\nabla_Y X,\cdot).
\tag{19.10}
$$

Then

$$
Q(Y,Y) \,=\, 2\int_0^{\overline{\tau}} \sqrt{\tau}\langle Y, TY\rangle\, d\tau \,+\, 2\,\sqrt{\overline{\tau}}\,\langle \nabla_X Y(\overline{\tau}), Y(\overline{\tau})\rangle.
\tag{19.11}
$$

An $\mathcal{L}$-Jacobi field along an $\mathcal{L}$-geodesic is a field $Y(\tau)$ that is annihilated by T.

The Hessian of the function $L(\cdot, \overline{\tau})$ can be computed as follows. Assume that $q \in M$ is outside of the time-$\overline{\tau}$ $\mathcal{L}$-cut locus. Let $\gamma : [0, \overline{\tau}] \rightarrow M$ be the minimizing $\mathcal{L}$-geodesic with $\gamma(0) = p$ and $\gamma(\overline{\tau}) = q$. Given $w \in T_q M$, take a short geodesic $c : (-\epsilon, \epsilon) \rightarrow M$ with $c(0) = q$ and $c'(0) = w$. Form the 1-parameter family of $\mathcal{L}$-geodesics $\tilde{\gamma}(s, \tau)$ with $\tilde{\gamma}(s, 0) = p$ and $\tilde{\gamma}(s, \overline{\tau}) = c(s)$. Then $Y(\tau) = \left.\frac{\partial \tilde{\gamma}(s,\tau)}{\partial s}\right|_{s=0}$ is an $\mathcal{L}$-Jacobi field Y along γ with $Y(0) = 0$ and $Y(\overline{\tau}) = w$. We have

$$\mathrm{Hess}_L(w, w) = \left.\frac{d^2 L(c(s))}{ds^2}\right|_{s=0} = Q(Y, Y) = 2\sqrt{\overline{\tau}}\, \langle \nabla_X Y(\overline{\tau}), Y(\overline{\tau}) \rangle. \tag{19.12}$$

From (19.11), a minimizer of $Q(Y, Y)$, among fields Y with given values at the endpoints, is an $\mathcal{L}$-Jacobi field. It follows that $\mathrm{Hess}_L(w, w) \leq Q(Y, Y)$ for *any* field Y along γ satisfying $Y(0) = 0$ and $Y(\overline{\tau}) = w$.

20. I.(7.8)-(7.9). Hessian bound for L

In this section we use a test variation field in order to estimate the Hessian of L.

If $Y(\overline{\tau})$ is a unit vector at $\gamma(\overline{\tau})$, solve for $\widetilde{Y}(\tau)$ in the equation

$$\nabla_X \widetilde{Y} = -\mathrm{Ric}(\widetilde{Y}, \cdot) + \frac{1}{2\tau} \widetilde{Y}, \tag{20.1}$$

on the interval $0 < \tau \leq \overline{\tau}$ with the endpoint condition $\widetilde{Y}(\overline{\tau}) = Y(\overline{\tau})$. (For this section, we change notation from the Y in I.7 to $\widetilde{Y}$.) Then

$$\frac{d}{d\tau} \langle \widetilde{Y}, \widetilde{Y} \rangle = 2\,\mathrm{Ric}(\widetilde{Y}, \widetilde{Y}) + 2 \langle \nabla_X \widetilde{Y}, \widetilde{Y} \rangle = \frac{1}{\tau} \langle \widetilde{Y}, \widetilde{Y} \rangle, \tag{20.2}$$

so $\langle \widetilde{Y}(\tau), \widetilde{Y}(\tau) \rangle = \frac{\tau}{\overline{\tau}}$. Thus we can extend $\widetilde{Y}$ continuously to the interval $[0, \overline{\tau}]$ by putting $\widetilde{Y}(0) = 0$. Substituting into (19.9) gives

(20.3)

$$\begin{aligned} Q(\widetilde{Y}, \widetilde{Y}) = & \int_0^{\overline{\tau}} \sqrt{\tau} \left[\mathrm{Hess}_R(\widetilde{Y}, \widetilde{Y}) + 2 \langle R(\widetilde{Y}, X) \widetilde{Y}, X \rangle + 2 \left| -\mathrm{Ric}(\widetilde{Y}, \cdot) + \frac{1}{2\tau} \widetilde{Y} \right|^2 \right. \\ & \qquad \left. - 4 (\nabla_{\widetilde{Y}} \mathrm{Ric})(\widetilde{Y}, X) + 2 (\nabla_X \mathrm{Ric})(\widetilde{Y}, \widetilde{Y}) \right] d\tau \\ = & \int_0^{\overline{\tau}} \sqrt{\tau} \left[\mathrm{Hess}_R(\widetilde{Y}, \widetilde{Y}) + 2 \langle R(\widetilde{Y}, X) \widetilde{Y}, X \rangle + 2 (\nabla_X \mathrm{Ric})(\widetilde{Y}, \widetilde{Y}) \right. \\ & \qquad \left. - 4 (\nabla_{\widetilde{Y}} \mathrm{Ric})(\widetilde{Y}, X) + 2 |\mathrm{Ric}(\widetilde{Y}, \cdot)|^2 - \frac{2}{\tau} \mathrm{Ric}(\widetilde{Y}, \widetilde{Y}) + \frac{1}{2\tau\overline{\tau}} \right] d\tau. \end{aligned}$$

From

(20.4)

$$\begin{aligned} \frac{d}{d\tau} \mathrm{Ric}(\widetilde{Y}(\tau), \widetilde{Y}(\tau)) = & \mathrm{Ric}_\tau(\widetilde{Y}, \widetilde{Y}) + (\nabla_X \mathrm{Ric})(\widetilde{Y}, \widetilde{Y}) + 2\,\mathrm{Ric}(\nabla_X \widetilde{Y}, \widetilde{Y}) \\ = & \mathrm{Ric}_\tau(\widetilde{Y}, \widetilde{Y}) + (\nabla_X \mathrm{Ric})(\widetilde{Y}, \widetilde{Y}) + \frac{1}{\tau} \mathrm{Ric}(\widetilde{Y}, \widetilde{Y}) - 2 |\mathrm{Ric}(\widetilde{Y}, \cdot)|^2, \end{aligned}$$

one obtains
(20.5)

$$- 2\sqrt{\overline{\tau}}\,\mathrm{Ric}(Y(\overline{\tau}), Y(\overline{\tau})) \;=\; -\,2\int_0^{\overline{\tau}} \frac{d}{d\tau}\left(\sqrt{\tau}\;\mathrm{Ric}(\widetilde{Y},\widetilde{Y})\right)\,d\tau \;=$$

$$-\int_0^{\overline{\tau}} \sqrt{\tau}\left[\frac{1}{\tau}\;\mathrm{Ric}(\widetilde{Y},\widetilde{Y}) \;+\; 2\,\mathrm{Ric}_\tau(\widetilde{Y},\widetilde{Y}) \;+\; 2(\nabla_X \mathrm{Ric})(\widetilde{Y},\widetilde{Y}) \;+\; \frac{2}{\tau}\;\mathrm{Ric}(\widetilde{Y},\widetilde{Y}) \;-\; 4|\,\mathrm{Ric}(\widetilde{Y},\cdot)|^2\right].$$

Combining (20.3) and (20.5) gives

$$\mathrm{Hess}_L(Y(\overline{\tau}),Y(\overline{\tau})) \,\le\, Q(\widetilde{Y},\widetilde{Y}) \,=\, \frac{1}{\sqrt{\overline{\tau}}} - 2\sqrt{\overline{\tau}}\,\mathrm{Ric}(Y(\overline{\tau}),Y(\overline{\tau})) - \int_0^{\overline{\tau}} \sqrt{\tau}\,H(X,\widetilde{Y})\,d\tau, \tag{20.6}$$

where
(20.7)

$$\begin{aligned} H(X,\widetilde{Y}) \;=\; &-\;\mathrm{Hess}_R(\widetilde{Y},\widetilde{Y}) \;-\; 2\langle R(\widetilde{Y},X)\widetilde{Y},X\rangle \;-\; 4\left(\nabla_X \mathrm{Ric}(\widetilde{Y},\widetilde{Y}) \;-\; \nabla_{\widetilde{Y}} \mathrm{Ric}(\widetilde{Y},X)\right) \\ &-\; 2\,\mathrm{Ric}_\tau(\widetilde{Y},\widetilde{Y}) \;+\; 2\big|\,\mathrm{Ric}(\widetilde{Y},\cdot)\big|^2 \;-\; \frac{1}{\tau}\;\mathrm{Ric}(\widetilde{Y},\widetilde{Y}). \end{aligned}$$

is the expression appearing in (F.4), after the change $\tau = -t$ and $X \to -X$. Note that $H(X,\widetilde{Y})$ is a quadratic form in $\widetilde{Y}$. For its relation to the expression $H(X)$ from (18.9), see Appendix F.

21. I.(7.10). The Laplacian of L

In this section we estimate $\triangle L$.

Let $\{Y_i(\overline{\tau})\}_{i=1}^n$ be an orthonormal basis of $T_{\gamma(\overline{\tau})}M$. Solve for $\widetilde{Y}_i(\tau)$ from (20.1). Putting $\widetilde{Y}_i(\tau) = \left(\frac{\tau}{\overline{\tau}}\right)^{1/2} e_i(\tau)$, the vectors $\{e_i(\tau)\}_{i=1}^n$ form an orthonormal basis of $T_{\gamma(\tau)}M$. Substituting into (20.6) and summing over i gives

$$\triangle L \;\le\; \frac{n}{\sqrt{\overline{\tau}}} \;-\; 2\sqrt{\overline{\tau}}R \;-\; \frac{1}{\overline{\tau}}\int_0^{\overline{\tau}} \tau^{3/2}\,\sum_i H(X,e_i)\,d\tau. \tag{21.1}$$

Then from (F.8),

$$\begin{aligned} \triangle L \;&\le \frac{n}{\sqrt{\overline{\tau}}} \;-\; 2\sqrt{\overline{\tau}}R \;-\; \frac{1}{\overline{\tau}}\int_0^{\overline{\tau}} \tau^{3/2}\,H(X)\,d\tau \\ &= \frac{n}{\sqrt{\overline{\tau}}} \;-\; 2\sqrt{\overline{\tau}}R \;-\; \frac{1}{\overline{\tau}}\,K. \end{aligned} \tag{21.2}$$

22. I.(7.11). Estimates on $\mathcal{L}$-Jacobi fields

In this section we estimate the growth rate of an $\mathcal{L}$-Jacobi field.

Given an $\mathcal{L}$-Jacobi field Y, we have

$$\frac{d}{d\tau}|Y|^2 \;=\; 2\,\mathrm{Ric}(Y,Y) \;+\; 2\langle\nabla_X Y,Y\rangle \;=\; 2\,\mathrm{Ric}(Y,Y) \;+\; 2\langle\nabla_Y X,Y\rangle. \tag{22.1}$$

Thus

$$\frac{d|Y|^2}{d\tau}\Big|_{\tau=\overline{\tau}} = 2\,\mathrm{Ric}(Y(\overline{\tau}), Y(\overline{\tau})) + \frac{1}{\sqrt{\overline{\tau}}}\,\mathrm{Hess}_L(Y(\overline{\tau}), Y(\overline{\tau})), \tag{22.2}$$

where we have used (19.12).

Let $\tilde{Y}$ be a field along γ as in Section 20, satisfying (20.1) with $\tilde{Y}(\overline{\tau}) = Y(\overline{\tau})$ and $|Y(\overline{\tau})| = 1$. Then from (20.6),

$$\frac{1}{\sqrt{\overline{\tau}}}\,\mathrm{Hess}_L(Y(\overline{\tau}), Y(\overline{\tau})) \le \frac{1}{\overline{\tau}} - 2\,\mathrm{Ric}(Y(\overline{\tau}), Y(\overline{\tau})) - \frac{1}{\sqrt{\overline{\tau}}}\int_0^{\overline{\tau}} \sqrt{\tau}\, H(X, \tilde{Y})\, d\tau. \tag{22.3}$$

Thus

$$\frac{d|Y|^2}{d\tau}\Big|_{\tau=\overline{\tau}} \le \frac{1}{\overline{\tau}} - \frac{1}{\sqrt{\overline{\tau}}}\int_0^{\overline{\tau}} \sqrt{\tau}\, H(X, \tilde{Y})\, d\tau. \tag{22.4}$$

The inequality is sharp if and only if the first inequality in (20.6) is an equality. This is the case if and only if $\tilde{Y}$ is actually the $\mathcal{L}$-Jacobi field Y, in which case

$$\frac{1}{\overline{\tau}} = \frac{d|\tilde{Y}|^2}{d\tau}\Big|_{\tau=\overline{\tau}} = \frac{d|Y|^2}{d\tau}\Big|_{\tau=\overline{\tau}} = 2\,\mathrm{Ric}(Y(\overline{\tau}), Y(\overline{\tau})) + \frac{1}{\sqrt{\overline{\tau}}}\,\mathrm{Hess}_L(Y(\overline{\tau}), Y(\overline{\tau})). \tag{22.5}$$

23. Monotonicity of the reduced volume $\widetilde{V}$

In this section we show that the reduced volume $\widetilde{V}(\tau)$ is monotonically nonincreasing in τ.

Fix $p \in M$. Define $l(q,\tau)$ as in (15.3). In order to show that $\tilde{V}(\tau)$ is well-defined in the noncompact case, we will need a lower bound on $l(q,\tau)$. For later use, we prove something slightly more general. Recall that we are assuming that we have bounded curvature on compact time intervals, and that time slices are complete.

Lemma 23.1. *Given $0 < \overline{\tau}_1 \le \overline{\tau}_2$, there constants $C_1, C_2 > 0$ so that for all $\tau \in [\overline{\tau}_1, \overline{\tau}_2]$ and all $q \in M$, we have*

$$l(q,\tau) \ge C_1\, d(p,q)^2 - C_2. \tag{23.2}$$

Proof. We write $\mathcal{L}$ in the form (17.6). Given an $\mathcal{L}$-geodesic γ with $\gamma(0) = p$ and $\gamma(\tau) = q$, we obtain

$$\mathcal{L}(\gamma) \ge \frac{1}{2}\int_0^{\sqrt{\tau}} \left|\frac{d\gamma}{ds}\right|^2 ds - \mathrm{const.} \tag{23.3}$$

As the multiplicative change in the metric between times $\overline{\tau}_1$ and $\overline{\tau}_2$ is bounded by a factor $e^{\mathrm{const.}\,(\overline{\tau}_2 - \overline{\tau}_1)}$, it follows that $L(q,\tau) \ge \mathrm{const.}\, d(p,q)^2 - \mathrm{const.}$, where the distance is measured at time τ. The lemma follows. □

Define $\tilde{V}(\tau)$ as in (15.4). As the volume of time-τ balls in M increases at most exponentially fast in the radius, it follows that $\tilde{V}(\tau)$ is well-defined. From the discussion in Section

17, we can write

$$\tilde{V}(\tau) \;=\; \int_{T_pM} \tau^{-\frac{n}{2}}\, e^{-l(\mathcal{L}exp_\tau(v),\tau)}\, \mathcal{J}(v,\tau)\, \chi_\tau(v)\, dv, \tag{23.4}$$

where $\mathcal{J}(v,\tau) \;=\; \det\, d\left(\mathcal{L}exp_\tau\right)_v$ is the Jacobian factor in the change of variable and χ_τ is the characteristic function of the time-τ domain Ω_τ of Section 17.

We first show that for each v, the expression $-\,\frac{n}{2}\ln(\tau) \,-\, l(\mathcal{L}exp_\tau(v),\tau) \,+\, \ln \mathcal{J}(v,\tau)$ is nonincreasing in τ. Let γ be the $\mathcal{L}$-geodesic with initial vector $v \in T_pM$. From (18.4) and (18.12),

$$\frac{dl(\gamma(\tau),\tau)}{d\tau}\Big|_{\tau=\overline{\tau}} \;=\; -\,\frac{1}{2\overline{\tau}}\, l(\gamma(\overline{\tau})) \;+\; \frac{1}{2}\,\left(R(\gamma(\overline{\tau})) \;+\; |X(\overline{\tau})|^2\right) \;=\; -\,\frac{1}{2}\,\overline{\tau}^{-\frac{3}{2}}\, K. \tag{23.5}$$

Next, let $\{Y_i\}_{i=1}^n$ be a basis for the Jacobi fields along γ that vanish at $\tau \;=\; 0$. We can write

$$\ln \mathcal{J}(v,\tau)^2 \;=\; \ln\det\,\left(\left(d\left(\mathcal{L}exp_\tau\right)_v\right)^*\, d\left(\mathcal{L}exp_\tau\right)_v\right) \;=\; \ln\det(S(\tau)) \;+\; \text{const.}, \tag{23.6}$$

where S is the matrix

$$S_{ij}(\tau) \;=\; \langle Y_i(\tau), Y_j(\tau)\rangle. \tag{23.7}$$

Then

$$\frac{d\ln\mathcal{J}(v,\tau)}{d\tau} \;=\; \frac{1}{2}\;\mathrm{Tr}\left(S^{-1}\frac{dS}{d\tau}\right). \tag{23.8}$$

To compute the derivative at $\tau \;=\; \overline{\tau}$, we can choose a basis so that $S(\overline{\tau}) \;=\; I_n$, i.e. $\langle Y_i(\overline{\tau}), Y_j(\overline{\tau})\rangle \;=\; \delta_{ij}$. Then using (22.4) and computing as in Section 21,

$$\frac{d\ln\mathcal{J}(v,\tau)}{d\tau}\Big|_{\tau=\overline{\tau}} \;=\; \frac{1}{2}\sum_{i=1}^n \frac{d|Y_i|^2}{d\tau}\Big|_{\tau=\overline{\tau}} \;\le\; \frac{n}{2\overline{\tau}} \;-\; \frac{1}{2}\,\overline{\tau}^{-\frac{3}{2}}K. \tag{23.9}$$

If we have equality then (22.5) holds for each Y_i, i.e. $2\,\mathrm{Ric} + \frac{1}{\sqrt{\overline{\tau}}}\;\mathrm{Hess}_L \;=\; \frac{g}{\overline{\tau}}$ at $\gamma(\overline{\tau})$.

From (23.5) and (23.9), we deduce that $\tau^{-\frac{n}{2}}\, e^{-l(\mathcal{L}exp_\tau(v),\tau)}\, \mathcal{J}(v,\tau)$ is nonincreasing in τ. Finally, recall that if $\tau \,\le\, \tau'$ then $\Omega_{\tau'} \subset \Omega_\tau$, so $\chi_\tau(v)$ is nonincreasing in τ. Hence $\tilde{V}(\tau)$ is nonincreasing in τ. If it is not strictly decreasing then we must have

$$2\,\mathrm{Ric}(\tau) \;+\; \frac{1}{\sqrt{\tau}}\;\mathrm{Hess}_{L(\tau)} \;=\; \frac{g(\tau)}{\tau}. \tag{23.10}$$

on all of M. Hence we have a gradient shrinking soliton solution.

24. I.(7.15). A differential inequality for L

In this section we discuss an important differential inequality concerning the reduced length l. We use the differential inequality to estimate $\min l(\cdot,\tau)$ from above. We then give a lower bound on l.

With $\overline{L}(q,\tau) \;=\; 2\sqrt{\tau}\, L(q,\tau)$, equations (18.13) and (21.2) imply that

$$\overline{L}_{\overline{\tau}} \;+\; \triangle\overline{L} \;\le\; 2n \tag{24.1}$$

away from the time-$\overline{\tau}$ $\mathcal{L}$-cut locus of p. We will eventually apply the maximum principle to this differential inequality. However to do so, we must discuss several senses in which the inequality can be made global on M, i.e. how it can be interpreted on the cut locus.

The first sense is that of a barrier differential inequality. Given $f \in C(M)$ and a function g on M, one says that $\triangle f \leq g$ *in the sense of barriers* if for all $q \in M$ and $\epsilon > 0$, there is a neighborhood V of q and some $u_\epsilon \in C^2(V)$ so that $u_\epsilon(q) = f(q)$, $u_\epsilon \geq f$ on V and $\triangle u_\epsilon \leq g + \epsilon$ on V [13]. There is a similar spacetime definition for $\triangle f - \frac{\partial f}{\partial t} \leq g$ [26]. The point of barrier differential inequalities is that they allow one to apply the maximum principle just as with smooth solutions.

We illustrate this by constructing a barrier function for $\overline{L}$ in (24.1). Given the spacetime point $(q, \overline{\tau})$, let $\gamma : [0, \overline{\tau}] \to M$ be a minimizing $\mathcal{L}$-geodesic with $\gamma(0) = p$ and $\gamma(\overline{\tau}) = q$. Given a small $\epsilon > 0$, let $u_\epsilon(q', \overline{\tau}')$ be the minimum of

$$\int_\epsilon^{\overline{\tau}'} \sqrt{\tau} \left(R(\gamma_2(\tau)) + \left|\dot{\gamma}_2(\tau)\right|^2 \right) d\tau + \int_0^\epsilon \sqrt{\tau} \left(R(\gamma(\tau)) + \left|\dot{\gamma}(\tau)\right|^2 \right) d\tau \tag{24.2}$$

among curves $\gamma_2 : [\epsilon, \overline{\tau}'] \to M$ with $\gamma_2(\epsilon) = \gamma(\epsilon)$ and $\gamma_2(\overline{\tau}') = q'$. Because the new basepoint $(\gamma(\epsilon), \epsilon)$ is moved in along γ from p, the minimizer γ_2 will be unique and will vary smoothly with q', when q' is close to q; otherwise a second minimizer or a "conjugate point" would imply that γ was not minimizing. Thus the function u_ϵ is smooth in a spacetime neighborhood V of $(q, \overline{\tau})$. Put $U_\epsilon(q', \overline{\tau}') = 2\sqrt{\overline{\tau}'}\, u_\epsilon(q', \overline{\tau}')$. By construction, $U_\epsilon \geq \overline{L}$ in V and $U_\epsilon(q, \overline{\tau}) = \overline{L}(q, \overline{\tau})$. For small ϵ, $(U_\epsilon)_{\overline{\tau}'} + \triangle U_\epsilon$ will be bounded above on V by something close to $2n$. Hence $\overline{L}$ satisfies (24.1) globally on M in the barrier sense.

As we are assuming bounded curvature on compact time intervals, we can now apply the maximum principle of Appendix A to conclude that the minimum of $\overline{L}(\cdot, \overline{\tau}) - 2n\overline{\tau}$ is nonincreasing in $\overline{\tau}$. (Note that from Lemma 23.1, the minimum of $\overline{L}(\cdot, \overline{\tau}) - 2n\overline{\tau}$ exists.)

Lemma 24.3. *For small positive $\overline{\tau}$, we have* $\min \overline{L}(\cdot, \overline{\tau}) - 2n\overline{\tau} < 0$.

Proof. Consider the static curve at the point p. Then for small $\overline{\tau}$, we have $\overline{L}(\cdot, \overline{\tau}) \leq \text{const.}\, \overline{\tau}^2$, from which the claim follows. ☐

(Being a bit more careful with the estimates in the proof of Lemma 23.1, one sees that $\lim_{\overline{\tau} \to 0} \min \overline{L}(\cdot, \overline{\tau}) = 0$.) Then for $\overline{\tau} > 0$, we must have $\min \overline{L}(\cdot, \overline{\tau}) \leq 2n\overline{\tau}$, so $\min l(\cdot, \overline{\tau}) \leq \frac{n}{2}$.

The other sense of a differential inequality is the distributional sense, i.e. $\triangle f \leq g$ if for every nonnegative compactly-supported smooth function ϕ on M,

$$\int_M (\triangle \phi)\, f\, dV \leq \int_M \phi\, g\, dV. \tag{24.4}$$

A general fact is that a barrier differential inequality implies a distributional differential inequality [37, 67].

We illustrate this by giving an alternative proof that $\tilde{V}(\tau)$ is nonincreasing in τ. From (18.13), (18.14) and (21.2), one finds that in the barrier sense (and hence in the distributional sense as well)

$$l_{\overline{\tau}} - \triangle l + |\nabla l|^2 - R + \frac{n}{2\overline{\tau}} \geq 0 \tag{24.5}$$

or, equivalently, that

$$(\partial_{\overline{\tau}} - \triangle)\left(\tau^{-\frac{n}{2}} e^{-l}\, dV\right) \leq 0. \tag{24.6}$$

Then for all nonnegative $\phi \in C_c^\infty(M)$ and $0 < \overline{\tau}_1 \leq \overline{\tau}_2$, one obtains

(24.7)

$$\int_M \phi\, \overline{\tau}_2^{-\frac{n}{2}}\, e^{-l(\cdot,\overline{\tau}_2)}\, dV(\overline{\tau}_2) - \int_M \phi\, \overline{\tau}_1^{-\frac{n}{2}}\, e^{-l(\cdot,\overline{\tau}_1)}\, dV(\overline{\tau}_1) =$$
$$\int_{\overline{\tau}_1}^{\overline{\tau}_2} \int_M \phi\, (\partial_{\overline{\tau}} - \triangle)\left(\tau^{-\frac{n}{2}} e^{-l(\cdot,\tau)}\, dV(\tau)\right) d\tau + \int_{\overline{\tau}_1}^{\overline{\tau}_2} \int_M (\triangle\phi)\, \tau^{-\frac{n}{2}}\, e^{-l(\cdot,\tau)}\, dV(\tau)\, d\tau \leq$$
$$\int_{\overline{\tau}_1}^{\overline{\tau}_2} \int_M (\triangle\phi)\, \tau^{-\frac{n}{2}}\, e^{-l(\cdot,\tau)}\, dV(\tau)\, d\tau.$$

We can find a sequence $\{\phi_i\}_{i=1}^\infty$ of such functions ϕ with range $[0,1]$ so that ϕ_i is one on $B(p,i)$, vanishes outside of $B(p,i^2)$, and $\sup_M |\triangle\phi_i| \leq i^{-1}$, uniformly in $\tau \in [\overline{\tau}_1, \overline{\tau}_2]$. Then to finish the argument it suffices to have a good upper bound on $e^{-l(\cdot,\tau)}$ in terms of $d(p,\cdot)$, uniformly in $\tau \in [\overline{\tau}_1, \overline{\tau}_2]$. This is given by Lemma 23.1. The monotonicity of $\tilde{V}$ follows.

We also note the equation

$$2\triangle l - |\nabla l|^2 + R + \frac{l-n}{\overline{\tau}} \leq 0, \tag{24.8}$$

which follows from (18.14) and (21.2).

Finally, suppose that the Ricci flow exists on a time interval $\tau \in [0,\tau_0]$. From the maximum principle of Appendix A, $R(\cdot,\tau) \geq -\frac{n}{2(\tau_0-\tau)}$. Then one obtains a lower bound on l as before, using the better lower bound for R.

25. I.7.2. Estimates on the reduced length

In this section we suppose that our solution has nonnegative curvature operator. We use this to derive estimates on the reduced length l.

We refer to Appendix F for Hamilton's differential Harnack inequality. We consider a Ricci flow defined on a time interval $t \in [0,\tau_0]$, with bounded nonnegative curvature operator, and put $\tau = \tau_0 - t$. The differential Harnack inequality gives the nonnegativity of the expression in (F.4). Comparing this with the formula for $H(X,Y)$ in (20.7), we can write the nonnegativity as

$$\left(H(X,Y) + \frac{\mathrm{Ric}(Y,Y)}{\tau}\right) + \frac{\mathrm{Ric}(Y,Y)}{\tau_0 - \tau} \geq 0, \tag{25.1}$$

or

$$H(X,Y) \geq -\mathrm{Ric}(Y,Y)\left(\frac{1}{\tau} + \frac{1}{\tau_0-\tau}\right). \tag{25.2}$$

Then

$$H(X) \geq -R\left(\frac{1}{\tau} + \frac{1}{\tau_0-\tau}\right). \tag{25.3}$$

As long as $\tau \le (1-c)\,\tau_0$, equations (18.3) and (18.12) give

$$
\begin{aligned}
(25.4) \qquad 4\tau|\nabla l|^2 &= -4\tau R + 4l - \frac{4}{\sqrt{\tau}} \int_0^\tau \tilde{\tau}^{3/2} H(X)\, d\tilde{\tau} \\
&\le -4\tau R + 4l + \frac{4}{\sqrt{\tau}} \int_0^\tau \tilde{\tau}^{3/2} R \left(\frac{1}{\tilde{\tau}} + \frac{1}{\tau_0 - \tilde{\tau}} \right) d\tilde{\tau} \\
&= -4\tau R + 4l + \frac{4}{\sqrt{\tau}} \int_0^\tau \sqrt{\tilde{\tau}} R \, \frac{\tau_0}{\tau_0 - \tilde{\tau}}\, d\tilde{\tau} \\
&\le -4\tau R + 4l + \frac{4}{c\sqrt{\tau}} \int_0^\tau \sqrt{\tilde{\tau}} R\, d\tilde{\tau} \\
&\le -4\tau R + 4l + \frac{8l}{c},
\end{aligned}
$$

where the last line uses (15.1). Thus

$$(25.5) \qquad |\nabla l|^2 + R \le \frac{Cl}{\tau}$$

for a constant $C = C(c)$. One shows similarly that

$$(25.6) \qquad \frac{d}{d\tau} \ln |Y|^2 \le \frac{1}{\tau}\,(Cl + 1)$$

for an $\mathcal{L}$-Jacobi field Y, using (22.4).

26. I.7.3. The no local collapsing theorem II

In this section we use the reduced volume to prove a no-local-collapsing theorem for a Ricci flow on a finite time interval.

Definition 26.1. We now say that a Ricci flow solution $g(\cdot)$ defined on a time interval $[0, T)$ is *κ-noncollapsed on the scale ρ* if for each $r < \rho$ and all $(x_0, t_0) \in M \times [0, T)$ with $t_0 \ge r^2$, whenever it is true that $|\operatorname{Rm}(x,t)| \le r^{-2}$ for every $x \in B_{t_0}(x_0, r)$ and $t \in [t_0 - r^2, t_0]$, then we also have $\operatorname{vol}(B_{t_0}(x_0, r)) \ge \kappa r^n$.

Definition 26.1 differs from Definition 13.14 by the requirement that the curvature bound holds in the entire parabolic region $B_{t_0}(x_0, r) \times [t_0 - r^2, t_0]$ instead of just on the ball $B_{t_0}(x_0, r)$ in the final time slice. Therefore a Ricci flow which is κ-noncollapsed in the sense of Definition 13.14 is also κ-noncollapsed in the sense of Definition 26.1.

Theorem 26.2. *Given numbers $n \in \mathbb{Z}^+$, $T < \infty$ and $\rho, K, c > 0$, there is a number $\kappa = \kappa(n, K, c, \rho, T) > 0$ with the following property. Let $(M^n, g(\cdot))$ be a Ricci flow solution defined on a time interval $[0, T)$ with $T < \infty$, such that the curvature $|\operatorname{Rm}|$ is bounded on every compact subinterval $[0, T'] \subset [0, T)$. Suppose that $(M, g(0))$ is a complete Riemannian manifold with $|\operatorname{Rm}| \le K$ and $\operatorname{inj}(M, g(0)) \ge c > 0$. Then the Ricci flow solution is κ-noncollapsed on the scale ρ, in the sense of Definition 26.1. Furthermore, with the other constants fixed, we can take κ to be nonincreasing in T.*

Proof. We first observe that the existence of $\mathcal{L}$-geodesics and the monotonicity of the reduced volume are valid in this setting; see Section 17.

Suppose that the theorem were false. Then for given $T < \infty$ and $\rho, K, c > 0$, there are :
1. A sequence $\{(M_k, g_k(\cdot))\}_{k=1}^{\infty}$ of Ricci flow solutions, each defined on the time interval $[0,T)$, with $|\operatorname{Rm}| \le K$ on $(M_k, g_k(0))$ and $\operatorname{inj}(M_k, g_k(0)) \ge c$,
2. Spacetime points $(p_k, t_k) \in M_k \times [0,T)$ and
3. Numbers $r_k \in (0, \rho)$
having the following property : $t_k \ge r_k^2$ and if we put $B_k = B_{t_k}(p_k, r_k) \subset M_k$ then $|\operatorname{Rm}|(x,t) \le r_k^{-2}$ whenever $x \in B_k$ and $t \in [t_k - r_k^2, t_k]$, but $\epsilon_k = r_k^{-1} \operatorname{vol}(B_k)^{\frac{1}{n}} \to 0$ as $k \to \infty$. From short-time curvature estimates along with the assumed bounded geometry at time zero, there is some $\overline{t} > 0$ so that we have uniformly bounded geometry on the time interval $[0, \overline{t}]$. In particular, we may assume that each t_k is greater than $\overline{t}$.

We define $\tilde{V}_k$ using curves going backward in real time from the basepoint (p_k, t_k), i.e. forward in τ-time from $\tau = 0$. The first step is to show that $\tilde{V}_k(\epsilon_k r_k^2)$ is small. Note that $\tau = \epsilon_k r_k^2$ corresponds to a real time of $t_k - \epsilon_k r_k^2$, which is very close to t_k.

Given an $\mathcal{L}$-geodesic $\gamma(\tau)$ with $\gamma(0) = p_k$ and velocity vector $X(\tau) = \frac{d\gamma}{d\tau}$, its initial vector is $v = \lim_{\tau \to 0} \sqrt{\tau}\, X(\tau) \in T_{p_k} M_k$. We first want to show that if $|v| \le .1\, \epsilon_k^{-1/2}$ then γ does not escape from B_k in time $\epsilon_k r_k^2$.

We have

$$\begin{aligned} \frac{d}{d\tau} \langle X(\tau), X(\tau) \rangle &= 2 \operatorname{Ric}(X,X) + 2 \langle X, \nabla_X X \rangle \\ &= 2 \operatorname{Ric}(X,X) + \langle X, \nabla R - \frac{1}{\tau} X - 4 \operatorname{Ric}(X, \cdot) \rangle \\ &= - \frac{|X|^2}{\tau} - 2 \operatorname{Ric}(X,X) + \langle X, \nabla R \rangle, \end{aligned} \tag{26.3}$$

so

$$\frac{d}{d\tau} \left(\tau |X|^2 \right) = -2\tau \operatorname{Ric}(X,X) + \tau \langle X, \nabla R \rangle. \tag{26.4}$$

Letting C denote a generic n-dependent constant, for $x \in B(p_k, r_k/2)$ and $t \in [t_k - r_k^2/2, t_k]$, the fact that g_k satisfies the Ricci flow gives an estimate $|\nabla R|(x,t) \le C r_k^{-3}$, as follows from the case $l = 0$, $m = 1$ of Appendix D. Then in terms of dimensionless variables,

$$\left| \frac{d}{d(\tau/r_k^2)} \left(\tau |X|^2 \right) \right| \le C\, \tau\, |X|^2 + C\, (\tau/r_k^2)^{1/2}\, (\tau |X|^2)^{1/2}. \tag{26.5}$$

Equivalently,

$$\left| \frac{d}{d(\tau/r_k^2)} \left(\sqrt{\tau}\, |X| \right) \right| \le C \sqrt{\tau}\, |X| + C\, \left(\tau/r_k^2 \right)^{1/2}. \tag{26.6}$$

Let us rewrite this as

$$\left| \frac{d}{d(\frac{\tau}{\epsilon_k r_k^2})} \left(\epsilon_k^{\frac{1}{2}} \sqrt{\tau}\, |X| \right) \right| \le C\, \epsilon_k \left(\epsilon_k^{\frac{1}{2}} \sqrt{\tau}\, |X| \right) + C\, \epsilon_k^2 \left(\frac{\tau}{\epsilon_k r_k^2} \right)^{1/2}. \tag{26.7}$$

We are interested in the time range when $\frac{\tau}{\epsilon_k r_k^2} \in [0,1]$ and the initial condition satisfies $\lim_{\tau \to 0} \epsilon_k^{\frac{1}{2}} \sqrt{\tau}\, |X|(\tau) \le .1$. Then because of the ϵ_k-factors on the right-hand side, it follows

from (26.7) that for large k, we will have $\epsilon_k^{\frac{1}{2}} \sqrt{\tau}\, |X|(\tau) \le .11$ for all $\tau \in [0, \epsilon_k r_k^2]$. Next,

$$(26.8) \quad \int_0^{\epsilon_k r_k^2} |X(\tau)|\, d\tau = \epsilon_k^{-\frac{1}{2}} \int_0^{\epsilon_k r_k^2} \epsilon_k^{\frac{1}{2}} \sqrt{\tau}\, |X|(\tau) \frac{d\tau}{\sqrt{\tau}} \le .11\, \epsilon_k^{-\frac{1}{2}} \int_0^{\epsilon_k r_k^2} \frac{d\tau}{\sqrt{\tau}} = .22\, r_k.$$

From the Ricci flow equation $g_\tau = 2\ \mathrm{Ric}$, it follows that the metrics $g(\tau)$ between $\tau = 0$ and $\overline{\tau} = \epsilon_k r_k^2$ are $e^{C\epsilon_k}$-biLipschitz close to each other. Then for ϵ_k small, the length of γ, as measured with the metric at time t_k, will be at most $.3\, r_k$. This shows that γ does not leave B_k within time $\epsilon_k r_k^2$.

Hence the contribution to $\tilde{V}_k(\epsilon_k r_k^2)$ coming from vectors $v \in T_{p_k} M_k$ with $|v| \le .1\, \epsilon_k^{-\frac{1}{2}}$ is at most $\int_{B_k} (\epsilon_k r_k^2)^{-\frac{n}{2}}\, e^{-\, l(q, \epsilon_k r_k^2)}\, dq$. We now want to give a lower bound on $l(q, \epsilon_k r_k^2)$ for $q \in B_k$. Given the $\mathcal{L}$-geodesic $\gamma : [0, \epsilon_k r_k^2] \to M_k$ with $\gamma(0) = p_k$ and $\gamma(\epsilon_k r_k^2) = q$, we have

$$(26.9) \quad \mathcal{L}(\gamma) \ge \int_0^{\epsilon_k r_k^2} \sqrt{\tau}\, R(\gamma(\tau))\, d\tau \ge - \int_0^{\epsilon_k r_k^2} \sqrt{\tau}\, n(n-1)\, r_k^{-2}\, d\tau = -\frac{2}{3}\, n(n-1)\, \epsilon_k^{\frac{3}{2}}\, r_k.$$

Then

$$(26.10) \qquad l(q, \epsilon_k r_k^2) \ge -\frac{1}{3}\, n(n-1)\, \epsilon_k.$$

Thus the contribution to $\tilde{V}_k(\epsilon_k r_k^2)$ coming from vectors $v \in T_{p_k} M_k$ with $|v| \le .1\, \epsilon_k^{-\frac{1}{2}}$ is at most

$$(26.11) \qquad e^{\frac{1}{3}\, n(n-1)\, \epsilon_k}\, (\epsilon_k r_k^2)^{-\frac{n}{2}}\ \mathrm{vol}_{t_k - \epsilon_k r_k^2}(B_k) \le e^{\frac{1}{3}\, n(n-1)\, \epsilon_k}\, e^{\mathrm{const.}\ \frac{1}{r_k^2}\, \epsilon_k r_k^2}\, \epsilon_k^{\frac{n}{2}},$$

which is less than $2\epsilon_k^{\frac{n}{2}}$ for large k.

To estimate the contribution to $\tilde{V}_k(\epsilon_k r_k^2)$ coming from vectors $v \in T_{p_k} M_k$ with $|v| > .1 \epsilon_k^{-\frac{1}{2}}$, we can use the previously-shown monotonicity of the integrand in τ. As $\tau \to 0$, the Euclidean calculation of Section 16 shows that $\tau^{-n/2}\, e^{-\, l(\mathcal{L}exp_\tau(v), \tau)}\, \mathcal{J}(v, \tau) \to 2^n\, e^{-\,|v|^2}$. Then for all $\tau > 0$ and all $v \in \Omega_\tau$,

$$(26.12) \qquad \tau^{-n/2}\, e^{-\, l(\mathcal{L}_\tau(v), \tau)}\, \mathcal{J}(v, \tau) \le 2^n\, e^{-\,|v|^2},$$

giving

$$(26.13) \quad \int_{T_{p_k} M_k - B(0, .1\, \epsilon_k^{-1/2})} \tau^{-n/2}\, e^{-\, l(\mathcal{L}_\tau(v), \tau)}\, J(v, \tau)\, \chi_\tau\, d^n v \le 2^n \int_{T_{p_k} M_k - B(0, .1\, \epsilon_k^{-1/2})} e^{-\,|v|^2}\, d^n v \le e^{-\frac{1}{10 \epsilon_k}}$$

for k large.

The conclusion is that $\lim_{k \to \infty} \tilde{V}_k(\epsilon_k r_k^2) = 0$. We now claim that there is a uniform positive lower bound on $\tilde{V}_k(t_k)$.

To estimate $\tilde{V}_k(t_k)$ (where $\tau = t_k$ corresponds to $t = 0$), we choose a point q_k at time $t = \frac{\overline{t}}{2}$, i.e. at $\tau = t_k - \frac{\overline{t}}{2}$, for which $l(q_k, t_k - \overline{t}/2) \le \frac{n}{2}$; see Section 24. Then we consider the concatenation of a fixed curve $\gamma_1^{(k)} : [0, t_k - \overline{t}/2] \to M_k$, having $\gamma_1^{(k)}(0) = p_k$ and $\gamma_1^{(k)}(t_k - \overline{t}/2) = q_k$, with a fan of curves $\gamma_2^{(k)} : [t_k - \overline{t}/2, t_k] \to M_k$ having $\gamma_2^{(k)}(t_k - \overline{t}/2) = q_k$. Because of the uniformly bounded geometry in the spacetime region with $t \in [0, \overline{t}/2]$, we

can get an upper bound in this way for $l(\cdot, t_k)$ in a region around q_k. Integrating $e^{-l(\cdot,t_k)}$, we get a positive lower bound on $\tilde{V}_k(t_k)$ that is uniform in k.

As $\epsilon_k r_k^2 \to 0$, the monotonicity of $\tilde{V}$ implies that $\tilde{V}_k(t_k) \leq \tilde{V}_k(\epsilon_k r_k^2)$ for large k, which is a contradiction. □

27. I.8.3. Length distortion estimates

The distortion of distances under Ricci flow can be estimated in terms of the Ricci tensor. We first mention a crude estimate.

Lemma 27.1. *If* $\mathrm{Ric} \leq (n-1)\,K$ *then for* $t_1 > t_0$*,*

$$\frac{\mathrm{dist}_{t_1}(x_0, x_1)}{\mathrm{dist}_{t_0}(x_0, x_1)} \geq e^{-(n-1)K(t_1 - t_0)}. \tag{27.2}$$

Proof. For any curve $\gamma : [0, a] \to M$, we have

$$\frac{d}{dt} L(\gamma) = \frac{d}{dt} \int_0^a \sqrt{\left\langle \frac{d\gamma}{ds}, \frac{d\gamma}{ds} \right\rangle}\, ds = - \int_0^a \mathrm{Ric}\left(\frac{d\gamma}{ds}, \frac{d\gamma}{ds} \right) \frac{ds}{\left| \frac{d\gamma}{ds} \right|} \geq -(n-1)\,K\, L(\gamma). \tag{27.3}$$

Integrating gives

$$\frac{L(\gamma)\big|_{t_1}}{L(\gamma)\big|_{t_0}} \geq e^{-(n-1)K(t_1 - t_0)}. \tag{27.4}$$

The lemma follows by taking γ to be a minimal geodesic at time t_1 between x_0 and x_1. □

Remark 27.5. By a similar argument, if $\mathrm{Ric} \geq -(n-1)\,K$ then for $t_1 > t_0$,

$$\frac{\mathrm{dist}_{t_1}(x_0, x_1)}{\mathrm{dist}_{t_0}(x_0, x_1)} \leq e^{(n-1)K(t_1 - t_0)}. \tag{27.6}$$

We can write the conclusion of Lemma 27.1 as

$$\frac{d}{dt} \mathrm{dist}_t(x_0, x_1) \geq -(n-1)K\ \mathrm{dist}_t(x_0, x_1), \tag{27.7}$$

where the derivative is interpreted in the sense of forward difference quotients.

The estimate in Lemma 27.1 is multiplicative. We now give an estimate that is additive in the distance.

Lemma 27.8. *(cf. Lemma I.8.3(b)) Suppose* $\mathrm{dist}_{t_0}(x_0, x_1) \geq 2r_0$*, and* $\mathrm{Ric}(x, t_0) \leq (n-1)K$ *for all* $x \in B_{t_0}(x_0, r_0) \cup B_{t_0}(x_1, r_0)$*. Then*

$$\frac{d}{dt} \mathrm{dist}_t(x_0, x_1) \geq -2\,(n-1) \left(\frac{2}{3} K\, r_0 + r_0^{-1} \right) \tag{27.9}$$

at time $t = t_0$*.*

Proof. If γ is a normalized minimal geodesic from x_0 to x_1 with velocity field $X(s) = \frac{d\gamma}{ds}$ then for any piecewise-smooth normal vector field V along γ that vanishes at the endpoints, the second variation formula gives

$$\int_0^{d(x_0,x_1)} \left(\left| \nabla_X V \right|^2 + \langle R(V,X)V, X \rangle \right) ds \geq 0. \tag{27.10}$$

Let $\{e_i(s)\}_{i=1}^{n-1}$ be a parallel orthonormal frame along γ that is perpendicular to X. Put $V_i(s) = f(s)\, e_i(s)$, where

$$f(s) = \begin{cases} \frac{s}{r_0} & \text{if } 0 \leq s \leq r_0, \\ 1 & \text{if } r_0 \leq s \leq d(x_0,x_1) - r_0, \\ \frac{d(x_0,x_1)-s}{r_0} & \text{if } d(x_0,x_1) - r_0 \leq s \leq d(x_0,x_1). \end{cases} \tag{27.11}$$

Then $\left| \nabla_X V_i \right| = |f'(s)|$ and

$$\int_0^{d(x_0,x_1)} \left| \nabla_X V_i \right|^2 ds = 2 \int_0^{r_0} \frac{1}{r_0^2}\, ds = \frac{2}{r_0}. \tag{27.12}$$

Next,

$$\begin{aligned} \int_0^{d(x_0,x_1)} \langle R(V_i,X)V_i, X \rangle\, ds = & \int_0^{r_0} \frac{s^2}{r_0^2} \langle R(e_i,X)e_i, X \rangle\, ds + \\ & \int_{r_0}^{d(x_0,x_1)-r_0} \langle R(e_i,X)e_i, X \rangle\, ds + \\ & \int_{d(x_0,x_1)-r_0}^{d(x_0,x_1)} \frac{(d(x_0,x_1)-s)^2}{r_0^2} \langle R(e_i,X)e_i, X \rangle\, ds. \end{aligned} \tag{27.13}$$

Then

$$\begin{aligned} 0 \leq & \sum_{i=1}^{n-1} \int_0^{d(x_0,x_1)} \left(\left| \nabla_X V_i \right|^2 + \langle R(V_i,X)V_i, X \rangle \right) ds \\ = & \frac{2(n-1)}{r_0} - \int_0^{d(x_0,x_1)} \operatorname{Ric}(X,X)\, ds + \int_0^{r_0} \left(1 - \frac{s^2}{r_0^2} \right) \operatorname{Ric}(X,X)\, ds + \\ & \int_{d(x_0,x_1)-r_0}^{d(x_0,x_1)} \left(1 - \frac{(d(x_0,x_1)-s)^2}{r_0^2} \right) \operatorname{Ric}(X,X)\, ds. \end{aligned} \tag{27.14}$$

This gives

$$\begin{aligned}
\frac{d}{dt}\,\mathrm{dist}_t(x_0,x_1) \;&=\; -\int_0^{d(x_0,x_1)} \mathrm{Ric}(X,X)\,ds \\
&\geq\; -\,\frac{2(n-1)}{r_0} \;-\; \int_0^{r_0}\left(1-\frac{s^2}{r_0^2}\right)\mathrm{Ric}(X,X)\,ds \\
&\quad\; -\int_{d(x_0,x_1)-r_0}^{d(x_0,x_1)}\left(1-\frac{(d(x_0,x_1)-s)^2}{r_0^2}\right)\mathrm{Ric}(X,X)\,ds \\
&\geq\; -\,\frac{2(n-1)}{r_0} \;-\; 2(n-1)K\cdot\frac{2}{3}\,r_0,
\end{aligned} \tag{27.15}$$

which proves the lemma. □

We now give an additive version of Lemma 27.1.

Corollary 27.16. [33, Theorem 17.2] *If* $\mathrm{Ric} \leq K$ *with* $K > 0$ *then for all* $x_0, x_1 \in M$*,*

$$\frac{d}{dt}\,\mathrm{dist}_t(x_0,x_1) \;\geq\; -\;\mathrm{const.}(n)\,K^{1/2}. \tag{27.17}$$

Proof. Put $r_0 = K^{-1/2}$. If $\mathrm{dist}_t(x_0,x_1) \leq 2r_0$ then the corollary follows from (27.7). If $\mathrm{dist}_t(x_0,x_1) > 2r_0$ then it follows from Lemma 27.8. □

The proof of the next lemma is similar to that of Lemma 27.8 and is given in I.8.

Lemma 27.18. *(cf. Lemma I.8.3(a)) Suppose that* $\mathrm{Ric}(x,t_0) \leq (n-1)\,K$ *on* $B_{t_0}(x_0,r_0)$*. Then the distance function* $d(x,t) = \mathrm{dist}_t(x,x_0)$ *satisfies*

$$d_t - \triangle d \;\geq\; -\,(n-1)\left(\frac{2}{3}\,K\,r_0 \;+\; r_0^{-1}\right) \tag{27.19}$$

at time $t = t_0$*, outside of* $B_{t_0}(x_0,r_0)$*. The inequality must be understood in the barrier sense (see Section 24) if necessary.*

28. I.8.2. No local collapsing propagates forward in time and to larger scales

This section is concerned with a localized version of the no-local-collapsing theorem. The main result, Theorem 28.2, says that noncollapsing propagates forward in time and to a larger distance scale.

We first give a local version of Definition 26.1.

Definition 28.1. (cf. Definition of I.8.1) A Ricci flow solution is said to be *κ-collapsed at* (x_0,t_0)*, on the scale* $r > 0$, if $|\mathrm{Rm}|(x,t) \leq r^{-2}$ for all $(x,t) \in B_{t_0}(x_0,r) \times [t_0-r^2,t_0]$, but $\mathrm{vol}(B_{t_0}(x_0,r^2)) \leq \kappa r^n$.

Theorem 28.2. *(cf. Theorem I.8.2) For any* $0 < A < \infty$*, there is some* $\kappa = \kappa(A) > 0$ *with the following property. Let* $g(\cdot)$ *be a Ricci flow solution defined for* $t \in [0, r_0^2]$*, having complete time slices and uniformly bounded sectional curvature.. Suppose that* $\mathrm{vol}(B_0(x_0,r_0)) \geq$

$A^{-1}r_0^n$ and that $|\operatorname{Rm}|(x,t) \le \frac{1}{nr_0^2}$ for all $(x,t) \in B_0(x_0,r_0)\times[0,r_0^2]$. Then the solution cannot be κ-collapsed on a scale less than r_0 at any point (x,r_0^2) with $x \in B_{r_0^2}(x_0,Ar_0)$.

Remark 28.3. In [51, Theorem I.8.2] the assumption is that $|\operatorname{Rm}|(x,t) \le r_0^{-2}$. We make the slightly stronger assumption that $|\operatorname{Rm}|(x,t) \le \frac{1}{nr_0^2}$. The extra factor of n is needed in order to assert in the proof that the region $\{(y,t) : \operatorname{dist}_{\frac{1}{2}}(y,x_0) \le \frac{1}{10},\ t \in [0,\frac{1}{2}]\}$ has bounded geometry; see below. Clearly this change of hypothesis does not make any substantial difference in the sequel.

Proof. We follow the lines of the proof of Theorem 26.2. By scaling, we can take $r_0 = 1$. Choose $x \in M$ with $\operatorname{dist}_1(x,x_0) < A$. Define the reduced volume $\tilde{V}(\tau)$ by means of curves starting at $(x,1)$. An effective lower bound on $\tilde{V}(1)$ would imply that the solution is not κ-collapsed at $(x,1)$, on a scale less than 1, for an appropriate $\kappa > 0$.

We first note that the geometry of the region $\{(y,t) : \operatorname{dist}_{\frac{1}{2}}(y,x_0) \le \frac{1}{10},\ t \in [0,\frac{1}{2}]\}$ is uniformly bounded. To see this, the upper sectional curvature bound implies that $\operatorname{Ric} \le 1$, so the distance distortion estimate of Section 27 implies that $B_{\frac{1}{2}}(x_0,\frac{1}{10}) \subset B_0(x_0,1)$. In particular, $|\operatorname{Rm}|(y,t) \le \frac{1}{n}$ on the region. By Remark 27.5, if $\operatorname{dist}_{\frac{1}{2}}(y,x_0) \le \frac{1}{10}$ and $t \in [0,\frac{1}{2}]$ then $B_0(y,\frac{1}{1000}) \subset B_t(y,\frac{1}{100})$. The Bishop-Gromov inequality gives a lower bound for the time-zero volume of $B_0(y,\frac{1}{1000})$, of the form $\operatorname{vol}_0(B_0(y,\frac{1}{1000})) \ge C_1(n,A)$. The Ricci flow equation then gives a lower bound for the time-t volume of $B_0(y,\frac{1}{1000})$, of the form $\operatorname{vol}_t(B_0(y,\frac{1}{1000})) \ge C_2(n,A)$. Thus the time-$t$ volume of $B_t(y,\frac{1}{100})$ satisfies $\operatorname{vol}(B_t(y,\frac{1}{100})) \ge C_2(n,A)$. This, along with the uniform sectional curvature bound, implies that the region has uniformly bounded geometry.

If we have an effective upper bound on $\min_y l(y,\frac{1}{2})$, where y ranges over points that satisfy $\operatorname{dist}_{\frac{1}{2}}(y,x_0) \le \frac{1}{10}$, then we obtain a lower bound on $\tilde{V}(1)$. Thus it suffices to obtain an effective upper bound on $\min_y l(y,\frac{1}{2})$ or, equivalently, on $\min_y \overline{L}(y,\frac{1}{2})$ (as defined using $\mathcal{L}$-geodesics from $(x,1)$) for y satisfying $\operatorname{dist}_{\frac{1}{2}}(y,x_0) \le \frac{1}{10}$. Applying the maximum principle to (24.1) gave an upper bound on $\inf_M \overline{L}$. The idea is to spatially localize this estimate near x_0, by means of a radial function ϕ.

Let $\phi = \phi(u)$ be a smooth function that equals 1 on $(-\infty,\frac{1}{20})$, equals infinity on $(\frac{1}{10},\infty)$ and is increasing on $(\frac{1}{20},\frac{1}{10})$, with

$$2(\phi')^2/\phi - \phi'' \ \ge\ (2A+100n)\phi' - C(A)\phi \tag{28.4}$$

for some constant $C(A) < \infty$. To satisfy (28.4), it suffices to take $\phi(u) = \frac{1}{e^{(2A+100n)(\frac{1}{10}-u)}-1}$ for u near $\frac{1}{10}$.

We claim that $\overline{L} + 2n + 1 \ge 1$ for $t \ge \frac{1}{2}$. To see this, from the end of Section 24,

$$R(\cdot,\tau) \ \ge\ -\frac{n}{2(1-\tau)}. \tag{28.5}$$

Then for $\overline{\tau} \in [0,\frac{1}{2}]$,

$$L(q,\overline{\tau}) \ \ge\ -\int_0^{\overline{\tau}} \sqrt{\tau}\,\frac{n}{2(1-\tau)}d\tau \ \ge\ -n\int_0^{\overline{\tau}} \sqrt{\tau}\,d\tau \ =\ -\frac{2n}{3}\,\overline{\tau}^{3/2}. \tag{28.6}$$

Hence

$$\overline{L}(q,\overline{\tau}) \;=\; 2\,\sqrt{\overline{\tau}}\,L(q,\overline{\tau}) \;\geq\; -\frac{4n}{3}\,\overline{\tau}^2 \;\geq\; -\,\frac{n}{3}, \tag{28.7}$$

which proves the claim.

Now put

$$h(y,t) \;=\; \phi(d(y,t)-A(2t-1))\,(\overline{L}(y,1-t)+2n+1), \tag{28.8}$$

where $d(y,t)=\mathrm{dist}_t(y,x_0)$. It follows from the above claim that $h(y,t)\;\geq\;0$ if $t\geq\frac{1}{2}$. Also,

$$\min_y h(y,1) \;\leq\; h(x,1) \;=\; \phi(\mathrm{dist}_1(x,x_0)-A)\cdot(2n+1) \;=\; 2n+1. \tag{28.9}$$

As ϕ is infinite on $(\frac{1}{10},\infty)$ and $\overline{L}(\cdot,\frac{1}{2})+2n+1\geq 1$, the minimum of $h(\cdot,\frac{1}{2})$ is achieved at some y satisfying $d(y,\frac{1}{2})\;\leq\;\frac{1}{10}$.

The calculations in I.8 give

$$\Box h \;\geq\; -(2n+C(A))h \tag{28.10}$$

at a minimum point of h, where $\Box \;=\; \partial_t-\triangle$. Then $\frac{d}{dt}h_{min}(t)\;\geq\;-(2n+C(A))\,h_{min}(t)$, so

$$h_{min}\left(\frac{1}{2}\right) \;\leq\; e^{n+\frac{C(A)}{2}}\,h_{min}(1) \;\leq\; (2n+1)\,e^{n+\frac{C(A)}{2}}. \tag{28.11}$$

It follows that

$$\min_{y\,:\,d(y,\frac{1}{2})\leq\frac{1}{10}} \overline{L}(y,\frac{1}{2})+2n+1 \;\leq\; (2n+1)\,e^{n+\frac{C(A)}{2}}. \tag{28.12}$$

This implies the theorem. □

29. I.9. Perelman's differential Harnack inequality

This section is concerned with a localized version of the $\mathcal{W}$-functional. It is mainly used in I.10.

Let $g(\cdot)$ be a Ricci flow solution on a manifold M, defined for $t\in(a,b)$. Put $\Box\;=\;\partial_t-\triangle$. For $f_1,f_2\in C_c^\infty((a,b)\times M)$, we have

$$\begin{aligned} 0 \;&=\; \int_a^b \frac{d}{dt}\int_M f_1(t,x)f_2(t,x)\,dV\,dt \\ &=\; \int_a^b\int_M ((\partial_t-\triangle)f_1)\,f_2\,dV \;+\; \int_a^b\int_M f_1\,(\partial_t+\triangle\;-\;R)f_2\,dV \\ &=\; \int_a^b\int_M (\Box f_1)\,f_2\,dV \;-\; \int_a^b\int_M f_1\,\Box^* f_2\,dV, \end{aligned} \tag{29.1}$$

where $\Box^*\;=\;-\,\partial_t-\triangle\;+\;R$. In this sense, $\Box^*$ is the formal adjoint to $\Box$.

Now suppose that the Ricci flow is defined for $t\in[0,T)$. Suppose that

$$u \;=\; (4\pi(T-t))^{-\frac{n}{2}}\,e^{-f} \tag{29.2}$$

satisfies $\Box^*u\;=\;0$. Put

$$v \;=\; [(T-t)(2\triangle f-|\nabla f|^2+R)+f-n]\,u. \tag{29.3}$$

If M is compact then using (5.10),

$$\mathcal{W}(g_{ij}, f, T-t) \ = \ \int_M v \, dV. \tag{29.4}$$

Proposition 29.5. *(cf. Proposition I.9.1)*

$$\Box^* v \ = \ -2(T-t) \left| R_{ij} + \nabla_i \nabla_j f - \frac{g_{ij}}{2(T-t)} \right|^2 u. \tag{29.6}$$

Proof. We note that the right-hand side of I.(9.1) should be multiplied by u.

To prove the proposition, we first claim that

$$\frac{d\triangle}{dt} \ = \ 2 \, R_{ij} \, \nabla_i \nabla_j. \tag{29.7}$$

To see this, for $f_1, f_2 \in C_c^\infty(M)$, we have

$$\int_M f_1 \triangle f_2 \, dV \ = \ - \int_M \langle df_1, df_2 \rangle \, dV. \tag{29.8}$$

Differentiating with respect to t gives

$$\int_M f_1 \frac{d\triangle}{dt} f_2 \, dV \ - \ \int_M f_1 \triangle f_2 \, R \, dV \ = \ -2 \int_M \mathrm{Ric}(df_1, df_2) \, dV \ + \ \int_M \langle df_1, df_2 \rangle \, R \, dV, \tag{29.9}$$

so

$$\frac{d\triangle}{dt} f_2 \ - \ R \triangle f_2 \ = \ 2 \, \nabla_i (R_{ij} \, \nabla_j f_2) \ - \ \nabla_i (R \, \nabla_i f_2). \tag{29.10}$$

Then (29.7) follows from the traced second Bianchi identity.

Next, one can check that $\Box^* u \ = \ 0$ is equivalent to

$$(\partial_t \ + \ \triangle) \, f \ = \ \frac{n}{2} \frac{1}{T-t} \ + \ |\nabla f|^2 \ - \ R. \tag{29.11}$$

Then one obtains

$$\begin{aligned} u^{-1} \, \Box^* v \ = \ & - \ (\partial_t \ + \ \triangle) \ \left[(T-t) \, (2\triangle f \ - \ |\nabla f|^2 \ + \ R) \ + \ f \right] \ - \\ & 2 \langle \nabla \left[(T-t) \, (2\triangle f \ - \ |\nabla f|^2 \ + \ R) \ + \ f \right], u^{-1} \nabla u \rangle \\ = \ & 2\triangle f \ - \ |\nabla f|^2 \ + \ R \ - \ (T-t) \, (\partial_t \ + \ \triangle) \, (2\triangle f \ - \ |\nabla f|^2 \ + \ R) \\ & - \ (\partial_t \ + \ \triangle) \, f \ + \ 2 \, (T-t) \, \langle \nabla (2\triangle f \ - \ |\nabla f|^2 \ + \ R), \nabla f \rangle \ + \ 2 \, |\nabla f|^2. \end{aligned} \tag{29.12}$$

Now

(29.13)

$$\begin{aligned} (\partial_t \ + \ \triangle) \, (2\triangle f \ - \ |\nabla f|^2 \ + \ R) \ = \ & 2(\partial_t \triangle) f \ + \ 2\triangle \, (\partial_t \ + \ \triangle) \, f \\ & - \ (\partial_t \ + \ \triangle) \, |\nabla f|^2 \ + \ (\partial_t \ + \ \triangle) \, R \\ = \ & 4 \, R_{ij} \, \nabla_i \nabla_j f \ + \ 2 \, \triangle \, (|\nabla f|^2 \ - \ R) \ - \ 2 \, \mathrm{Ric}(df, df) \\ & - \ 2 \langle \nabla f_t, \nabla f \rangle \ - \ \triangle |\nabla f|^2 \ + \ \triangle R \ + \ 2 \, |\mathrm{Ric}|^2 \ + \ \triangle R \\ = \ & 4 \, R_{ij} \, \nabla_i \nabla_j f \ + \ 2 \, \triangle \, |\nabla f|^2 \ - \ 2 \, \mathrm{Ric}(df, df) \\ & - \ 2 \langle \nabla (- \triangle f \ + \ |\nabla f|^2 \ - \ R), \nabla f \rangle \ - \ \triangle |\nabla f|^2 \ + \ 2 \, |\mathrm{Ric}|^2. \end{aligned}$$

Hence the term in $u^{-1} \Box^* v$ proportionate to $(T-t)^{-1}$ is

$$- \frac{n}{2} \frac{1}{T-t}. \tag{29.14}$$

The term proportionate to $(T-t)^0$ is

$$2\triangle f - |\nabla f|^2 + R - |\nabla f|^2 + R + 2|\nabla f|^2 = 2(\triangle f + R). \tag{29.15}$$

The term proportionate to $(T-t)$ is $(T-t)$ times

$$\begin{aligned} &- 4 R_{ij} \nabla_i \nabla_j f - 2 \triangle |\nabla f|^2 + 2 \operatorname{Ric}(df, df) + \\ &2\langle \nabla(- \triangle f + |\nabla f|^2 - R), \nabla f \rangle + \triangle |\nabla f|^2 - 2 |\operatorname{Ric}|^2 + \\ &2 \langle \nabla(2\triangle f - |\nabla f|^2 + R), \nabla f \rangle = \\ &- 4 R_{ij} \nabla_i \nabla_j f - \triangle |\nabla f|^2 + 2 \operatorname{Ric}(df, df) + 2\langle \nabla \triangle f, \nabla f \rangle - 2 |\operatorname{Ric}|^2 = \\ &- 4 R_{ij} \nabla_i \nabla_j f - 2 |\operatorname{Hess}(f)|^2 - 2 |\operatorname{Ric}|^2. \end{aligned} \tag{29.16}$$

Putting this together gives

$$\Box^* v = - 2 (T-t) \left| R_{ij} + \nabla_i \nabla_j f - \frac{1}{2(T-t)} g_{ij} \right|^2 u. \tag{29.17}$$

This proves the proposition. $\Box$

As a consequence of Proposition 29.5,

$$\begin{aligned} \frac{d}{dt} \mathcal{W}(g_{ij}, f, T-t) &= \frac{d}{dt} \int_M v \, dV = \int_M (\partial_t + \triangle - R) v \, dV \\ &= 2 (T-t) \int_M \left| R_{ij} + \nabla_i \nabla_j f - \frac{1}{2(T-t)} g_{ij} \right|^2 u \, dV. \end{aligned} \tag{29.18}$$

In this sense, Proposition 29.5 is a local version of the monotonicity of $\mathcal{W}$.

Corollary 29.19. *(cf. Corollary I.9.2) If M is closed, or whenever the maximum principle holds, then $\max v/u$ is nondecreasing in t.*

Proof. We note that the statement of Corollary I.9.2 should have $\max v/u$ instead of $\min v/u$.

To prove the corollary, we have

$$(\partial_t + \triangle) \frac{v}{u} = \frac{v \Box^* u - u \Box^* v}{u^2} - \frac{2}{u} \left\langle \nabla u, \nabla \frac{v}{u} \right\rangle. \tag{29.20}$$

As $\Box^* u = 0$ and $\Box^* v \leq 0$, the corollary now follows from the maximum principle. $\Box$

We now assume that the Ricci flow solution is defined on the closed interval $[0, T]$.

Corollary 29.21. *(cf. Corollary I.9.3) Under the same assumptions, if the solution is defined for $t \in [0, T]$ and u tends to a δ-function as $t \to T$ then $v \leq 0$ for all $t < T$.*

Proof. Suppose that h is a positive solution of $\Box h = 0$. Then

$$\frac{d}{dt} \int_M hv \, dV = \int_M ((\Box h) v - h \Box^* v) \, dV = - \int_M h \Box^* v \, dV \geq 0. \tag{29.22}$$

As $t \to T$, the computation of $\int_M hv$ approaches the flat-space calculation, which one finds to be zero; see [50] for details. (Strictly speaking, the paper [50] deals with the case when M is closed. It is indicated that the proof should extend to the noncompact setting.) Thus $\int_M h(t_0) v(t_0)\, dV$ is nonpositive for all $t_0 < T$. As $h(t_0)$ can be taken to be an arbitrary positive function, and then flowed forward to a positive solution of $\Box h = 0$, it follows that $v(t_0) \le 0$ for all $t_0 < T$. □

The next result compares the function f used in the $\mathcal{W}$-functional and the function l used in the reduced volume.

Corollary 29.23. *(cf. Corollary I.9.5) Under the assumptions of the previous corollary, let $p \in M$ be the point where the limit δ-function is concentrated. Then $f(q,t) \le l(q, T-t)$, where l is the reduced distance defined using curves starting from (p,T).*

Proof. Equation (24.6) implies that $\Box^* \left((4\pi\tau)^{-n/2} \, e^{-l} \right) \le 0$. (This corrects the statement at the top of page 23 of I.) From this and the fact that $\Box^* \left((4\pi\tau)^{-n/2} \, e^{-f} \right) = 0$, the argument of the proof of Corollary 29.19 gives that $\max e^{f-l}$ is nondecreasing in t, so $\max(f-l)$ is nondecreasing in t. As $t \to T$ one obtains the flat-space result, namely that $f - l$ vanishes. Thus $f(t) \le l(T-t)$ for all $t \in [0,T)$. □

Remark 29.24. To give an alternative proof of Corollary 29.23, putting $\tau = T - t$, Corollary I.9.4 of [51] says that for any smooth curve γ,

$$\frac{d}{d\tau} f(\gamma(\tau),\tau) \le \frac{1}{2} \left(R(\gamma(\tau),\tau) + \left| \dot\gamma(t) \right|^2 \right) - \frac{1}{2\tau} f(\gamma(\tau),\tau), \tag{29.25}$$

or

$$\frac{d}{d\tau} \left(\tau^{1/2} f(\gamma(\tau),\tau) \right) \le \frac{1}{2} \tau^{1/2} \left(R(\gamma(\tau),\tau) + \left| \dot\gamma(t) \right|^2 \right). \tag{29.26}$$

Take γ to be a curve emanating from (p,T). For small τ,

$$f(\gamma(\tau),\tau) \sim d(p,\gamma(\tau))^2/4\tau = O(\tau^0). \tag{29.27}$$

Then integration gives $\tau^{1/2} f \le \frac{1}{2} L$, or $f \le l$.

30. The statement of the pseudolocality theorem

The next theorem says that, in a localized sense, if the initial data of a Ricci flow solution has a lower bound on the scalar curvature and satisfies an isoperimetric inequality close to that of Euclidean space then there is a sectional curvature bound in a forward region. The result is not used in the sequel.

Theorem 30.1. *(cf. Theorem I.10.1) For every $\alpha > 0$ there exist $\delta, \epsilon > 0$ with the following property. Suppose that we have a smooth pointed Ricci flow solution $(M, (x_0, 0), g(\cdot))$ defined for $t \in [0, (\epsilon r_0)^2]$, such that each time slice is complete. Suppose that for any $x \in B_0(x_0, r_0)$ and $\Omega \subset B_0(x_0, r_0)$, we have $R(x, 0) \ge -r_0^{-2}$ and $\mathrm{vol}(\partial\Omega)^n \ge (1-\delta)\, c_n \, \mathrm{vol}(\Omega)^{n-1}$, where c_n is the Euclidean isoperimetric constant. Then $|\operatorname{Rm}|(x,t) < \alpha t^{-1} + (\epsilon r_0)^{-2}$ whenever $0 < t \le (\epsilon r_0)^2$ and $d(x,t) = \mathrm{dist}_t(x, x_0) \le \epsilon r_0$.*

The sectional curvature bound $|\operatorname{Rm}|(x,t) < \alpha t^{-1} + (\epsilon r_0)^{-2}$ necessarily blows up as $t \to 0$, as nothing was assumed about the sectional curvature at $t = 0$.

We first sketch the idea of the proof of Theorem 30.1. It is an argument by contradiction. One takes a Ricci flow solution that satisfies the assumptions and picks a point $(\overline{x}, \overline{t})$ where the desired curvature bound does not hold. One can assume, roughly speaking, that $(\overline{x}, \overline{t})$ is the first point in the given solution where the bound does not hold. (This will give the curvature bound needed for taking a limit in a sequence of counterexamples.) One now considers the solution u to the conjugate heat equation, starting as a δ-function at $(\overline{x}, \overline{t})$, and the corresponding function v. We know that $v \le 0$. The first goal is to get a negative upper bound for the integral of v over an appropriate ball B at a time $\widetilde{t}$ near $\overline{t}$; see Section 33. The argument to get such a bound is by contradiction. If there were not such a bound then one could consider a rescaled sequence of counterexamples with $\int_B v \, dV \to 0$, and try to take a limit. If one has the injectivity radius bounds needed to take a limit then one obtains a limit solution with $\int_B v \, dV = 0$, which implies that the limit solution is a gradient shrinking soliton, which violates curvature assumptions. If one doesn't have the injectivity radius bounds then one can do a further rescaling to see that in fact $\int_B v \, dV \to -\infty$ for some subsequence, which is a contradiction.

If M is compact then $\int_M v \, dV$ is monotonically nondecreasing in t. As (29.6) is a localized version of this statement, whether M is compact or noncompact we can use a cutoff function h and equation (29.6) to get a negative upper bound on $\int_M hv \, dV$ at time $t = 0$. Finally, $\int_M v \, dV$ is the expression that appears in the logarithmic Sobolev inequality. If the isoperimetric constant is sufficiently close to the Euclidean value c_n then one concludes that $\int_M hv \, dV$ must be bounded below by a constant close to zero, which contradicts the negative upper bound on $\int_M hv \, dV$.

31. Claim 1 of I.10.1. A point selection argument

In Theorem 30.1, we can assume that $r_0 = 1$ and $\alpha < \frac{1}{100n}$. Fix α and put $M_\alpha = \{(x,t) : |\operatorname{Rm}(x,t)| \ge \alpha t^{-1}\}$.

The next lemma says that if we have a point (x,t) where the conclusion of Theorem 30.1 does not hold then there is another point $(\overline{x}, \overline{t})$ with $|\operatorname{Rm}(\overline{x}, \overline{t})|$ large (relative to $\overline{t}^{-1}$) so that any other such point (x', t') either has $t' > \overline{t}$ or is much farther from x_0 than $\overline{x}$ is.

Lemma 31.1. *(cf. Claim 1 of I.10.1) For any $A > 0$, if $g(\cdot)$ is a Ricci flow solution for $t \in [0, \epsilon^2]$, with $A\epsilon < \frac{1}{100n}$, and $|\operatorname{Rm}|(x,t) \ge \alpha t^{-1} + \epsilon^{-2}$ for some (x,t) satisfying $t \in (0, \epsilon^2]$ and $d(x,t) \le \epsilon$, then one can find $(\overline{x}, \overline{t}) \in M_\alpha$ with $\overline{t} \in (0, \epsilon^2]$ and $d(\overline{x}, \overline{t}) < (2A+1)\epsilon$, such that*

$$|\operatorname{Rm}(x', t')| \le 4\, |\operatorname{Rm}(\overline{x}, \overline{t})| \tag{31.2}$$

whenever

$$(x', t') \in M_\alpha, \quad t' \in (0, \overline{t}], \quad d(x', t') \le d(\overline{x}, \overline{t}) + A |\operatorname{Rm}|^{-\frac{1}{2}}(\overline{x}, \overline{t}). \tag{31.3}$$

Proof. The proof is by a point selection argument as in Appendix H. By assumption, there is a point (x,t) satisfying $t \in (0, \epsilon^2]$, $d(x,t) \le \epsilon$ and $|\operatorname{Rm}(x,t)| \ge \alpha t^{-1} + \epsilon^{-2}$. Clearly $(x,t) \in M_\alpha$. Define points (x_k, t_k) inductively as follows. First, $(x_1, t_1) = (x,t)$. Next,

suppose that (x_k, t_k) is constructed but cannot be taken for $(\overline{x}, \overline{t})$. Then there is some point $(x_{k+1}, t_{k+1}) \in M_\alpha$ such that $0 < t_{k+1} \leq t_k$, $d(x_{k+1}, t_{k+1}) \leq d(x_k, t_k) + A|\operatorname{Rm}|^{-\frac{1}{2}}(x_k, t_k)$ and $|\operatorname{Rm}|(x_{k+1}, t_{k+1}) > 4|\operatorname{Rm}|(x_k, t_k)$. Continuing in this way, the point (x_k, t_k) constructed has $|\operatorname{Rm}|(x_k, t_k) \geq 4^{k-1}|\operatorname{Rm}|(x_1, t_1) \geq 4^{k-1}\epsilon^{-2}$. Then

$$\begin{aligned} d(x_k, t_k) &\leq d(x_1, t_1) + A|\operatorname{Rm}|^{-\frac{1}{2}}(x_1, t_1) + \ldots + A|\operatorname{Rm}|^{-\frac{1}{2}}(x_{k-1}, t_{k-1}) \\ &\leq \epsilon + 2A|\operatorname{Rm}|^{-\frac{1}{2}}(x_1, t_1) \leq (2A+1)\epsilon. \end{aligned} \tag{31.4}$$

As the solution is smooth, the induction process must terminate after a finite number of steps and the last value (x_k, t_k) can be taken for $(\overline{x}, \overline{t})$. □

32. Claim 2 of I.10.1. Getting parabolic regions

In Lemma 31.1, we know that (31.2) is satisfied under the condition (31.3). The spacetime region described in (31.3) is not a product region, due to the fact that $d(x, t)$ is time-dependent. The next goal is to obtain the estimate (31.2) on a product region in spacetime; this will be necessary when taking limits of Ricci flow solutions. To get the estimate on a product region, one needs to bound how fast distances are changing with respect to t.

Lemma 32.1. *(cf. Claim 2 of I.10.1) For the point $(\overline{x}, \overline{t})$ constructed in Lemma 31.1,*

$$|\operatorname{Rm}(x', t')| \;\leq\; 4\,|\operatorname{Rm}(\overline{x}, \overline{t})| \tag{32.2}$$

holds whenever

$$\overline{t} \;-\; \frac{1}{2}\,\alpha Q^{-1} \;\leq\; t' \;\leq\; \overline{t}, \quad \operatorname{dist}_{\overline{t}}(x', \overline{x}) \;\leq\; \frac{1}{10}\,AQ^{-\frac{1}{2}}, \tag{32.3}$$

where $Q = |\operatorname{Rm}(\overline{x}, \overline{t})|$.

Proof. We first claim that if (x', t') satisfies $\overline{t} - \frac{1}{2}\alpha Q^{-1} \leq t' \leq \overline{t}$ and $d(x', t') \leq d(\overline{x}, \overline{t}) + AQ^{-1/2}$ then $|\operatorname{Rm}|(x', t') \leq 4Q$. To see this, if $(x', t') \in M_\alpha$ then it is true by Lemma 31.1. If $(x', t') \notin M_\alpha$ then $|\operatorname{Rm}|(x', t') < \alpha(t')^{-1}$. As $(\overline{x}, \overline{t}) \in M_\alpha$, we know that $Q \geq \alpha\overline{t}^{-1}$. Then $t' \geq \overline{t} - \frac{1}{2}\,\alpha Q^{-1} \geq \frac{1}{2}\,\overline{t}$ and so $|\operatorname{Rm}|(x', t') < 2\,\alpha\overline{t}^{-1} \leq 2Q$.

Thus we have a uniform curvature bound on the time-t' distance ball $B(x_0, d(\overline{x}, \overline{t}) + AQ^{-1/2})$, provided that $\overline{t} - \frac{1}{2}\,\alpha Q^{-1} \leq t' \leq \overline{t}$. We now claim that the time-$\overline{t}$ ball $B(x_0, d(\overline{x}, \overline{t}) + \frac{1}{10}\,AQ^{-1/2})$ lies in the time-t' distance ball $B(x_0, d(\overline{x}, \overline{t}) + AQ^{-1/2})$. To see this, applying Lemma 27.8 with $r_0 = \frac{1}{100}AQ^{-1/2}$ and the above curvature bound, if x' is in the time-$\overline{t}$ ball $B(x_0, d(\overline{x}, \overline{t}) + \frac{1}{10}\,AQ^{-1/2})$ then

$$\operatorname{dist}_{t'}(x_0, x') \;-\; \operatorname{dist}_{\overline{t}}(x_0, x') \;\leq\; \frac{1}{2}\,\alpha Q^{-1} \,\cdot\, 2(n-1)\,\left(\frac{2}{3}\cdot 4Q(\frac{1}{100}AQ^{-1/2}) \;+\; 100\,A^{-1}Q^{1/2}\right). \tag{32.4}$$

Assuming that A is sufficiently large (we'll take $A \to \infty$ later) and using the fact that $\alpha < \frac{1}{100n}$, it follows that $d(x', t') \leq d(x', \overline{t}) + \frac{1}{2}AQ^{-1/2} \leq d(\overline{x}, \overline{t}) + AQ^{-\frac{1}{2}}$, which is what we want to show. We note that the argument also shows that is indeed self-consistent to use the curvature bounds in the application of Lemma 27.8.

Now suppose that (x', t') satisfies (32.3). By the triangle inequality, x' lies in the time-$\overline{t}$ distance ball $B(x_0, d(\overline{x}, \overline{t}) + \frac{1}{10} AQ^{-1/2})$. Then x' is in the time-t' distance ball $B(x_0, d(\overline{x}, \overline{t}) + AQ^{-1/2})$ and so $|\operatorname{Rm}(x', t')| \leq 4Q$, which proves the lemma. $\square$

33. Claim 3 of I.10.1. An upper bound on the integral of v

We first make some remarks about the fundamental solution to the backward heat equation. Let $(M, (\overline{x}, b), g(\cdot))$ be a smooth one-parameter family of complete pointed Riemannian manifolds, parametrized by $t \in (a, b]$. The fundamental solution u of the backward heat equation is a positive solution of $\square^* u = 0$ on $M \times (a, b)$ such that $u(\cdot, t)$ converges to $\delta_{\overline{x}}$ in the distributional sense, as $t \to b^-$. It is constructed as follows (cf. [26, Section 3]). Let $\{D_i\}_{i=1}^\infty$ be an exhaustion of M by an increasing sequence of smooth compact codimension-zero submanifolds-with-boundary containing $\overline{x}$ in the interior. Let $u^{(i)}$ be the unique solution of $\square^* u^{(i)} = 0$ on $D_i \times (a, b)$ with $\lim_{t \to b^-} u^{(i)}(x, t) = \delta_{\overline{x}}(x)$, as constructed using Dirichlet boundary conditions on D_i. If $D_i \subset D_j$ then $u^{(i)} \leq u^{(j)}$ on D_i. Then the fundamental solution is defined to be the limit $u = \lim_{i \to \infty} u^{(i)}$, with smooth convergence on compact subsets of $M \times (a, b)$. The function u is independent of the choice of exhaustion sequence $\{D_i\}_{i=1}^\infty$. For any $t \in (a, b)$, we have $\int_M u(x, t)\, dV(x) \leq 1$. If $\int_M u(x, t)\, dV(x) = 1$ for all t then we say that $(M, (\overline{x}, b), g(\cdot))$ is stochastically complete for $\square^*$. This will be the case if one has bounded curvature on compact time intervals, but need not be the case in general.

Lemma 33.1. *Let $\{(M_k, (\overline{x}_k, b), g_k(\cdot))\}_{k=1}^\infty$ be a sequence of manifolds as above, each defined on the time interval $(a, b]$. Suppose that $\lim_{k \to \infty} (M_k, (\overline{x}_k, b), g_k(\cdot)) = (M_\infty, (\overline{x}_\infty, b), g_\infty(\cdot))$ in the pointed smooth topology, and that $(M_\infty, (\overline{x}_\infty, b), g_\infty(\cdot))$ is stochastically complete for $\square^*$. Then after passing to a subsequence, the fundamental solutions $\{u_k\}_{k=1}^\infty$ converge smoothly on compact subsets of $M_\infty \times (a, b)$ to the fundamental solution u_∞. (Of course, we use the pointed diffeomorphisms inherent in the statement of pointed convergence in order to compare the u_k's with u_∞.)*

Proof. From the uniform upper L^1-bound on $\{u_k(\cdot, t)\}_{k=1}^\infty$ and parabolic regularity, after passing to a subsequence we can assume that $\{u_k\}_{k=1}^\infty$ converges smoothly on compact subsets of $M_\infty \times (a, b)$ to some function U. From the construction of u_∞, it follows easily that $u_\infty \leq U$. For any $t \in (a, b)$, we have

(33.2)
$$\begin{aligned} \int_{M_\infty} (U(x,t) - u_\infty(x,t)) \operatorname{dvol}(x) &= \int_{M_\infty} \liminf_k (u_k(x,t) - u_\infty(x,t)) \operatorname{dvol}(x) \\ &\leq \liminf_k \int_{M_\infty} (u_k(x,t) - u_\infty(x,t)) \operatorname{dvol}(x) \leq 0, \end{aligned}$$

so $U = u_\infty$. $\square$

Starting the proof of Theorem 30.1, we suppose that the theorem is not true. Then there are sequences $\epsilon_k \to 0$ and $\delta_k \to 0$, and pointed Ricci flow solutions $(M_k, (x_{0,k}, 0), g_k(\cdot))$ which satisfy the hypotheses of the theorem but for which there is a point (x_k, t_k) with $0 < t_k \leq \epsilon_k^2$, $d(x_k, t_k) \leq \epsilon_k$ and $|\operatorname{Rm}|(x_k, t_k) \geq \alpha t_k^{-1} + \epsilon_k^{-2}$. Given the flow $(M_k, g_k(\cdot))$, we reduce ϵ_k as

much as possible so that there is still such a point (x_k, t_k). Then

$$|\operatorname{Rm}|(x,t) < \alpha t_k^{-1} + 2\epsilon_k^{-2} \tag{33.3}$$

whenever $0 < t \le \epsilon_k^2$ and $d(x,t) \le \epsilon_k$. Put $A_k = \frac{1}{100n\epsilon_k}$. Construct points $(\overline{x}_k, \overline{t}_k)$ as in Lemma 31.1. Consider fundamental solutions $u_k = (4\pi(\overline{t}_k - t))^{-\frac{n}{2}} e^{-f_k}$ of $\square^* u_k = 0$ satisfying $\lim_{t \to \overline{t}_k^-} u(x,t) = \delta_{\overline{x}_k}(x)$. Construct the corresponding functions v_k from (29.3).

Lemma 33.4. *(cf. Claim 3 of I.10.1) There is some $\beta > 0$ so that for all sufficiently large k, there is some $\widetilde{t}_k \in [\overline{t}_k - \frac{1}{2}\alpha Q_k^{-1}, \overline{t}_k]$ with $\int_{B_k} v_k \, dV_k \le -\beta$, where $Q_k = |\operatorname{Rm}|(\overline{x}_k, \overline{t}_k)$ and B_k is the time-$\widetilde{t}_k$ ball of radius $\sqrt{\overline{t}_k - \widetilde{t}_k}$ centered at $\overline{x}_k$.*

Proof. Suppose that the claim is not true. After passing to a subsequence, we can assume that for any choice of $\widetilde{t}_k$, $\liminf_{k\to\infty} \int_{B_k} v_k \, dV_k \ge 0$.

Consider the pointed solution $(M_k, (\overline{x}_k, \overline{t}_k), g_k(\cdot))$ parabolically rescaled by Q_k. Suppose first that there is a subsequence so that the injectivity radii of the scaled metrics at $(\overline{x}_k, \overline{t}_k)$ are bounded away from zero. Since $A_k \to \infty$, we can use Lemma 32.1 and Appendix E to take a subsequence that converges to a complete Ricci flow solution $(M_\infty, (\overline{x}_\infty, \overline{t}_\infty), g_\infty(\cdot))$ on a time interval $(\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty]$, with $|\operatorname{Rm}| \le 4$ and $|\operatorname{Rm}|(\overline{x}_\infty, \overline{t}_\infty) = 1$. Consider the fundamental solution u_∞ of $\square^*$ on M_∞ with $\lim_{t\to\overline{t}_\infty^-} u_\infty(x_\infty, t) = \delta_{\overline{x}_\infty}(x_\infty)$. As before, let u_k be the fundamental solution of $\square^*$ on M_k with $\lim_{t\to\overline{t}_k^-} u_k(x_k, t) = \delta_{\overline{x}_k}(x_k)$. In view of the pointed convergence of the rescalings of $(M_k, (\overline{x}_k, \overline{t}_k), g_k(\cdot))$ to $(M_\infty, (\overline{x}_\infty, \overline{t}_\infty), g_\infty(\cdot))$, Lemma 33.1 implies that after passing to a further subsequence we can ensure that $\lim_{k\to\infty} u_k = u_\infty$, with smooth convergence on compact subsets of $M_\infty \times (\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty)$. (The curvature bounds on $(M_\infty, (\overline{x}_\infty, \overline{t}_\infty), g_\infty(\cdot))$ ensure that it is stochastically complete for $\square^*$.) From Corollary 29.21, $v_\infty \le 0$. Note that we are applying Corollary 29.21 on $M_\infty \times (\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty)$, where we have the curvature bounds needed to use the maximum principle.

Given $\widetilde{t}_\infty \in (\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty)$, let B_∞ be the time-$\widetilde{t}_\infty$ ball of radius $\sqrt{\overline{t}_\infty - \widetilde{t}_\infty}$ centered at $\overline{x}_\infty$. In view of the smooth convergence $\lim_{k\to\infty} u_k = u_\infty$ on compact subsets of $M_\infty \times (\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty)$, it follows that $\int_{B_\infty} v_\infty \, dV_\infty = 0$ at time $\widetilde{t}_\infty$, so v_∞ vanishes on B_∞ at time $\widetilde{t}_\infty$. Let h be a solution to $\square h = 0$ on $M_\infty \times [\widetilde{t}_\infty, \overline{t}_\infty)$ with $h(\cdot, \widetilde{t}_\infty)$ a nonnegative nonzero function supported in B_∞. As in the proof of Corollary 29.21, $\int_{M_\infty} h v_\infty \, dV_\infty$ is nondecreasing in t and vanishes for $t = \widetilde{t}_\infty$ and $t \to \overline{t}_\infty$. Thus $\int_{M_\infty} h v_\infty \, dV_\infty$ vanishes for all $t \in [\widetilde{t}_\infty, \overline{t}_\infty)$. However, for $t \in (\widetilde{t}_\infty, \overline{t}_\infty)$, h is strictly positive and v_∞ is nonpositive. Thus v_∞ vanishes on M_∞ for all $t \in (\widetilde{t}_\infty, \overline{t}_\infty)$, and so

$$\operatorname{Ric}(g_\infty) \, + \, \operatorname{Hess} f_\infty \, - \, \frac{1}{2(\overline{t} - t)} \, g_\infty \, = \, 0. \tag{33.5}$$

on this interval. We know that $|\operatorname{Rm}| \le 4$ on $M_\infty \times (\overline{t}_\infty - \frac{1}{2}\alpha, \overline{t}_\infty]$. From the evolution equation,

$$\frac{dg_\infty}{dt} \, = \, -2\operatorname{Ric}(g_\infty) \, = \, 2\operatorname{Hess} f_\infty \, - \, \frac{1}{\overline{t} - t} \, g_\infty. \tag{33.6}$$

It follows that the supremal and infimal sectional curvatures of $g_\infty(\cdot, t)$ go like $(\overline{t}_\infty - t)^{-1}$. Hence g_∞ is flat, which contradicts the fact that $|\operatorname{Rm}|(\overline{x}_\infty, \overline{t}_\infty) = 1$.

Suppose now that there is a subsequence so that the injectivity radii of the scaled metrics at $(\overline{x}_k, \overline{t}_k)$ tend to zero. Parabolically rescale $(M_k, (\overline{x}_k, \overline{t}_k), g_k(\cdot))$ further so that the injectivity radius becomes one. After passing to a subsequence we will have convergence to a flat Ricci flow solution $(-\infty, 0] \times L$. The complete flat manifold L can be described as the total space of a flat orthogonal $\mathbb{R}^m$-bundle over a flat compact manifold C. After separating variables, the fundamental solution u_∞ on L will be Gaussian in the fiber directions and will decay exponentially fast to a constant in the base directions, i.e. $u(x,\tau) \sim (4\pi\tau)^{-m/2}\, e^{-\frac{|x|^2}{4\tau}}\, \frac{1}{\mathrm{vol}(C)}$, where $|x|$ is the fiber norm. With this for u_∞, one finds that $v_\infty = (m-n)\left(1 + \frac{1}{2}\ln(4\pi\tau)\right) u_\infty$. With B_τ the ball around a basepoint of radius $\sqrt{\tau}$, the integral of u over B_τ has a positive limit as $\tau \to \infty$, and so $\lim_{\tau\to\infty} \int_{B_\tau} v_\infty \, dV_\infty = -\infty$. Then there are times $\widetilde{t}_k \in [\overline{t}_k - \frac{1}{2}\alpha Q_k^{-1}, \overline{t}_k]$ so that $\lim_{k\to\infty} \int_{B_k} v_k \, dV_k = -\infty$, which is a contradiction. □

34. Theorem I.10.1. Proof of the pseudolocality theorem

Continuing with the proof of Theorem 30.1, we now use Lemma 33.4 to get a contradiction to a log Sobolev inequality. For simplicity of notation, we drop the subscript k and deal with a particular $(M_k, (\overline{x}_k, \overline{t}_k), g_k(\cdot))$ for k large. Define a smooth function ϕ on $\mathbb{R}$ which is one on $(-\infty, 1]$, decreasing on $[1,2]$ and zero on $[2,\infty]$, with $\phi'' \geq -10\phi'$ and $(\phi')^2 \leq 10\phi$. To construct ϕ we can take the function which is 1 on $(-\infty, 1]$, $1 - 2(x-1)^2$ on $[1, 3/2]$, $2(x-2)^2$ on $[3/2, 2]$ and 0 on $[2,\infty)$, and smooth it slightly.

Put $\widetilde{d}(y,t) = d(y,t) + 200n\sqrt{t}$. We claim that if $10A\epsilon \leq \widetilde{d}(y,t) \leq 20A\epsilon$ then $d_t(y,t) - \triangle d(y,t) + \frac{100n}{\sqrt{t}} \geq 0$. To see this, recalling that $t \in [0, \epsilon^2]$, if $10A\epsilon \leq \widetilde{d}(y,t) \leq 20A\epsilon$ and A is sufficiently large then $9A\epsilon \leq d(y,t) \leq 21A\epsilon$. We apply Lemma 27.18 with the parameter r_0 of Lemma 27.18 equal to $\sqrt{t}$. As $r_0 \leq \epsilon$, we have $y \notin B(x_0, r_0)$. From (33.3), on $B(x_0, r_0)$ we have $|\operatorname{Rm}|(\cdot, t) \leq \alpha\, t^{-1} + 2\,\epsilon^{-2}$. Then from Lemma 27.18, at (y,t) we have

$$
\begin{aligned}
d_t - \triangle d &\geq -(n-1)\left(\frac{2}{3}(\alpha\, t^{-1} + 2\,\epsilon^{-2})t^{1/2} + t^{-1/2}\right) \\
&= -(n-1)\left(1 + \frac{2}{3}\alpha + \frac{4}{3}\,\epsilon^{-2} t\right) t^{-1/2}.
\end{aligned} \tag{34.1}
$$

It follows that $d_t - \triangle d + \frac{100n}{\sqrt{t}} \geq 0$.

Now put $h(y,t) = \phi\left(\frac{\widetilde{d}(y,t)}{10A\epsilon}\right)$. Then $\Box h = \frac{1}{10A\epsilon}\left(d_t - \triangle d + \frac{100n}{\sqrt{t}}\right)\phi' - \frac{1}{(10A\epsilon)^2}\phi''$, where the arguments of ϕ' and ϕ'' are $\frac{\widetilde{d}(y,t)}{10A\epsilon}$. Where $\phi' \neq 0$, we have $d_t - \triangle d + \frac{100n}{\sqrt{t}} \geq 0$. The fundamental solution $u(x,t) = (4\pi(\overline{t}-t))^{-\frac{n}{2}}\, e^{-f(x,t)}$ of $\Box^*$ is positive for $t \in [0, \overline{t})$ and we have $\int_M u \, dV \leq 1$ for all t. (Recall that we are not assuming stochastic completeness.)

Then

(34.2)
$$\left(\int_M hu\,dV\right)_t = \int_M ((\Box h)u - h\Box^* u)\,dV = \int_M (\Box h)u\,dV \le -\frac{1}{(10A\epsilon)^2}\int_M \phi'' u\,dV$$
$$\le \frac{10}{(10A\epsilon)^2}\int_M \phi u\,dV \le \frac{10}{(10A\epsilon)^2}\int_M u\,dV \le \frac{10}{(10A\epsilon)^2}.$$

Hence

$$\int_M hu\,dV\Big|_{t=0} \ge \int_M hu\,dV\Big|_{t=\bar{t}} - \frac{\bar{t}}{(A\epsilon)^2} \ge 1 - A^{-2}. \tag{34.3}$$

Similarly, using Proposition 29.5 and Corollary 29.21,

(34.4)
$$\left(-\int_M hv\,dV\right)_t = -\int_M ((\Box h)v - h\Box^* v)\,dV \le -\int_M (\Box h)v\,dV$$
$$\le \frac{1}{(10A\epsilon)^2}\int_M \phi'' v\,dV \le -\frac{10}{(10A\epsilon)^2}\int_M \phi v\,dV = -\frac{10}{(10A\epsilon)^2}\int_M hv\,dV.$$

Consider the time $\widetilde{t}$ of Lemma 33.4. As $(\overline{x}, \overline{t}) \in M_\alpha$, $\widetilde{t} \in [\overline{t}/2, \overline{t}]$. Then $\sqrt{\overline{t} - \widetilde{t}} \le 2^{-1/2}\,\epsilon$ and so for large A, h will be one on the ball B at time $\widetilde{t}$ of radius $\sqrt{\overline{t} - \widetilde{t}}$ centered at $\overline{x}$. Then at time $\widetilde{t}$,

$$-\int_M hv\,dV \ge -\int_B v\,dV \ge \beta. \tag{34.5}$$

Thus

$$-\int_M hv\,dV\Big|_{t=0} \ge \beta\, e^{-\frac{\widetilde{t}}{(A\epsilon)^2}} \ge \beta\, e^{-\frac{\overline{t}}{(A\epsilon)^2}} \ge \beta\left(1 - \frac{\overline{t}}{(A\epsilon)^2}\right) \ge \beta(1 - A^{-2}). \tag{34.6}$$

Working at time 0, put $\widetilde{u} = hu$ and $\widetilde{f} = f - \log h$. In what follows we implicitly integrate over $\mathrm{supp}(h)$. We have

$$\beta(1 - A^{-2}) \le -\int_M hv\,dV = \int_M [(-2\triangle f + |\nabla f|^2 - R)\overline{t} - f + n]hu\,dV. \tag{34.7}$$

We claim that

$$\int_M \left(-2\triangle f + |\nabla f|^2\right) he^{-f}\,dV = \int_M \left(-|\nabla \widetilde{f}|^2 + \frac{|\nabla h|^2}{h^2}\right) he^{-f}\,dV. \tag{34.8}$$

This follows from

$$
\begin{aligned}
(34.9)\qquad \int_M \left(-2\triangle f \,+\, |\nabla f|^2\right)\, he^{-f}\, dV \;&=\; \int_M \left(2\langle \nabla f, \nabla(he^{-f})\rangle \,+\, |\nabla f|^2\, he^{-f}\right)\, dV \\
&= \int_M \left(2\langle \nabla f, \frac{\nabla h}{h} - \nabla f\rangle \,+\, |\nabla f|^2\right)\, he^{-f}\, dV \\
&= \int_M \langle \nabla f, 2\frac{\nabla h}{h} - \nabla f\rangle\, he^{-f}\, dV \\
&= \int_M \langle \nabla \widetilde{f} + \frac{\nabla h}{h}, \frac{\nabla h}{h} - \nabla \widetilde{f}\rangle\, he^{-f}\, dV \\
&= \int_M \left(-\, |\nabla \widetilde{f}|^2 \,+\, \frac{|\nabla h|^2}{h^2}\right)\, he^{-f}\, dV.
\end{aligned}
$$

Then

$$
\begin{aligned}
(34.10)\qquad & \int_M [(-2\triangle f + |\nabla f|^2 - R)\overline{t} - f + n] hu\, dV \;= \\
& \int_M [-\overline{t}|\nabla \widetilde{f}|^2 - \widetilde{f} + n]\widetilde{u}\, dV \,+\, \int_M [\overline{t}(|\nabla h|^2/h - Rh) - h\log h] u\, dV.
\end{aligned}
$$

Next, $\frac{|\nabla h|^2}{h} \,\le\, \frac{10}{(10A\epsilon)^2}$ and $-\,Rh \,\le\, 1$ (from the assumed lower bound on R at time zero). Then

$$
(34.11)\qquad \int_M \overline{t}\, \left(\frac{|\nabla h|^2}{h} \,-\, Rh\right)\, u\, dV \,\le\, \epsilon^2\, \left(\frac{10}{(10A\epsilon)^2} \,+\, 1\right) \,\le\, A^{-2} \,+\, \epsilon^2.
$$

Also,

$$
\begin{aligned}
(34.12)\qquad -\int_M uh\log h\, dV \;&=\; -\int_{B(x_0,20A\epsilon)-B(x_0,10A\epsilon)} uh\log h\, dV \;\le\; \int_{M-B(x_0,10A\epsilon)} u\, dV \\
&\le\; 1 \,-\, \int_{B(x_0,10A\epsilon)} u\, dV.
\end{aligned}
$$

Putting $\overline{h}(y) = \phi\left(\frac{d(y)}{5A\epsilon}\right)$, a result similar to (34.3) shows that

$$
(34.13)\qquad \int_{B(x_0,10A\epsilon)} u\, dV \;\ge\; \int_M \overline{h} u\, dV \;\ge\; 1 \,-\, cA^{-2}
$$

for an appropriate constant c. Putting this together gives

$$
(34.14)\qquad \beta(1 - A^{-2}) \;\le\; \int_M \left(-\overline{t}|\nabla \widetilde{f}|^2 - \widetilde{f} + n\right) \widetilde{u}\, dV \,+\, (1+c)A^{-2} + \epsilon^2.
$$

Put $\widehat{g} = \frac{1}{2\overline{t}} g$, $\widehat{u} = (2\overline{t})^{\frac{n}{2}} \widetilde{u}$ and define $\widehat{f}$ by $\widehat{u} \,=\, (2\pi)^{-\frac{n}{2}}\, e^{-\widehat{f}}$. From (34.3) and (34.14), if we restore the subscript k then $\lim_{k\to\infty} \int_M \widehat{u}_k\, d\widehat{V}_k \,=\, 1$ and for large k,

$$
(34.15)\qquad \frac{1}{2}\,\beta \;\le\; \int_{M_k} \left(-\,\frac{1}{2}\,|\nabla \widehat{f}_k|^2 - \widehat{f}_k + n\right) \widehat{u}_k\, d\widehat{V}_k.
$$

If we normalize $\widehat{u}_k$ by putting $U_k = \frac{\widehat{u}_k}{\int_{M_k} \widehat{u}_k \, d\widehat{V}_k}$, and define F_k by $U_k = (2\pi)^{-\frac{n}{2}} e^{-F_k}$, then for large k, we also have

$$\frac{1}{2}\beta \leq \int_{M_k} \left(-\frac{1}{2}|\nabla F_k|^2 - F_k + n\right) U_k \, d\widehat{V}_k. \tag{34.16}$$

On the other hand, the logarithmic Sobolev inequality for $\mathbb{R}^n$ [8, I.(8)] says that

$$\int_{\mathbb{R}^n} \left(-\frac{1}{2}|\nabla F|^2 - F + n\right) U \, dV \leq 0, \tag{34.17}$$

provided that the compactly-supported function $U = (2\pi)^{-n/2} e^{-F}$ satisfies $\int_{\mathbb{R}^n} U \, dV = 1$. As was mentioned to us by Peter Topping, one can get a sharper inequality by applying (34.17) to the rescaled function $U_c(x) = c^n U(cx)$ and optimizing with respect to c. The result is

$$\int_{\mathbb{R}^n} |\nabla F|^2 U \, dV \geq n \, e^{1 - \frac{2}{n} \int_{\mathbb{R}^n} F \, U \, dV}. \tag{34.18}$$

Given this inequality on $\mathbb{R}^n$, one can use a symmetrization argument to prove the same inequality for a compactly-supported function on any complete Riemannian manifold, provided that the Euclidean isoperimetric inequality holds for domains in the support of U. See, for example, [48, Proposition 4.1] which gives the symmetrization argument for (34.17), attributing it to Perelman. Again using the inequality for $\mathbb{R}^n$, if instead we have $\mathrm{vol}(\partial\Omega)^n \geq (1-\delta_k) \, c_n \, \mathrm{vol}(\Omega)^{n-1}$ for domains $\Omega \subset \mathrm{supp}(U_k)$ then the symmetrization argument gives

$$\int_{M_k} |\nabla F_k|^2 \, U_k \, d\widehat{V}_k \geq (1-\delta_k)^{\frac{2}{n}} \, n \, e^{1 - \frac{2}{n} \int_{M_k} F_k \, U_k \, d\widehat{V}_k}. \tag{34.19}$$

Equations (34.16) and (34.19) imply that

$$\frac{n}{2} \left((1-\delta_k)^{\frac{2}{n}} \, e^{1 - \frac{2}{n} \int_{M_k} F_k \, U_k \, d\widehat{V}_k} - 1 - \left(1 - \frac{2}{n} \int_{M_k} F_k \, U_k \, d\widehat{V}_k\right)\right) \leq -\frac{\beta}{2}. \tag{34.20}$$

However,

$$\lim_{k \to \infty} \inf_{x \in \mathbb{R}} \left((1-\delta_k)^{\frac{2}{n}} \, e^x - 1 - x \right) = 0. \tag{34.21}$$

This is a contradiction.

35. I.10.2. The volumes of future balls

The next result gives a lower bound on the volumes of future balls.

Corollary 35.1. *(cf. Corollary I.10.2) Under the assumptions of Theorem 30.1, for* $0 < t \leq (\epsilon r_0)^2$ *we have* $\mathrm{vol}(B_t(x, \sqrt{t})) \geq c t^{\frac{n}{2}}$ *for* $x \in B_0(x_0, \epsilon r_0)$*, where* $c = c(n)$ *is a universal constant.*

Proof. (Sketch) If the corollary were not true then taking a sequence of counterexamples, we can center ourselves around the collapsing balls $B(x, \sqrt{t})$ to obtain functions f as in Section 34. As in the proof of Theorem 13.3, the volume condition along with the fact that $\int_M (2\pi)^{-n/2} e^{-f} dV \to 1$ means that $f \to -\infty$, which implies that $\int_M \left(- \frac{1}{2} |\nabla f|^2 - f + n \right) u dV \to \infty$. This contradicts the logarithmic Sobolev inequality. □

36. I.10.4. κ-noncollapsing at future times

The next result gives κ-noncollapsing at future times.

Corollary 36.1. *(cf. Corollary I.10.4) There are $\delta, \epsilon > 0$ such that for any $A > 0$ there exists $\kappa = \kappa(A) > 0$ with the following property. Suppose that we have a Ricci flow solution $g(\cdot)$ defined for $t \in [0, (\epsilon r_0)^2]$ which has bounded $|\operatorname{Rm}|$ and complete time slices. Suppose that for any $x \in B(x_0, r_0)$ and $\Omega \subset B(x_0, r_0)$, we have $R(x, 0) \geq -r_0^{-2}$ and $\operatorname{vol}(\partial\Omega)^n \geq (1 - \delta)\, c_n \operatorname{vol}(\Omega)^{n-1}$, where c_n is the Euclidean isoperimetric constant. If (x, t) satisfies $A^{-1}(\epsilon r_0)^2 \leq t \leq (\epsilon r_0)^2$ and $\operatorname{dist}_t(x, x_0) \leq A r_0$ then $g(\cdot)$ is not κ-collapsed at (x, t) on scales less than $\sqrt{t}$.*

Proof. Using Theorem 30.1 and Corollary 35.1, we can apply Theorem 28.2 starting at time $A^{-1}(\epsilon r_0)^2$. □

37. I.10.5. Diffeomorphism finiteness

In this section we prove the diffeomorphism finiteness of Riemannian manifolds with local isoperimetric inequalities, a lower bound on scalar curvature and an upper bound on volume.

Theorem 37.1. *Given $n \in \mathbb{Z}^+$, there is a $\delta > 0$ with the following property. For any $r_0, V > 0$, there are finitely many diffeomorphism types of compact n-dimensional Riemannian manifolds (M, g_0) satisfying*
1. $R \geq -r_0^{-2}$.
2. $\operatorname{vol}(M, g_0) \leq V$.
3. Any domain $\Omega \subset M$ contained in a metric r_0-ball satisfies $\operatorname{vol}(\partial\Omega)^n \geq (1-\delta) c_n \operatorname{vol}(\Omega)^{n-1}$, where c_n is the Euclidean isoperimetric constant.

Proof. Choose $\alpha > 0$. Let δ and ϵ be the parameters of Theorem 30.1. Consider Ricci flow $g(\cdot)$ starting from (M, g_0). Let $T > 0$ be the maximal number so that a smooth flow exists for $t \in [0, T)$. If $T < \infty$ then $\lim_{t \to T^-} \sup_{x \in M} |\operatorname{Rm}(x, t)| = \infty$. It follows from Theorem 30.1 that $T > (\epsilon r_0)^2$. Put $\widehat{g} = g((\epsilon r_0)^2)$. Theorem 30.1 gives a uniform double-sided sectional curvature bound on $(M, \widehat{g})$. Corollary 35.1 gives a uniform lower bound on the volumes of (ϵr_0)-balls in $(M, \widehat{g})$. Let $\{x_i\}_{i=1}^N$ be a maximal $(2\epsilon r_0)$-separated net in $(M, \widehat{g})$.

From the lower bound $R \geq -r_0^{-2}$ on (M, g_0) and the maximum principle, we have $R(x, t) \geq -r_0^{-2}$ for $t \in [0, (\epsilon r_0)^2]$. Then the Ricci flow equation gives a uniform upper bound on $\operatorname{vol}(M, \widehat{g})$. This implies a uniform upper bound on N or, equivalently, a uniform upper bound on $\operatorname{diam}(M, \widehat{g})$. The theorem now follows from the diffeomorphism finiteness of n-dimensional Riemannian manifolds with double-sided sectional curvature bounds, upper bounds on diameter and lower bounds on volume. □

38. I.11.1. κ-SOLUTIONS

Definition 38.1. Given $\kappa > 0$, a *κ-solution* is a Ricci flow solution $(M, g(\cdot))$ that is defined on a time interval of the form $(-\infty, C)$ (or $(-\infty, C]$) such that

- The curvature $|\,\mathrm{Rm}\,|$ is bounded on each compact time interval $[t_1, t_2] \subset (-\infty, C)$ (or $(-\infty, C]$), and each time slice $(M, g(t))$ is complete.
- The curvature operator is nonnegative and the scalar curvature is everywhere positive.
- The Ricci flow is κ-noncollapsed at all scales.

By abuse of terminology, we may sometimes write that "$(M, g(\cdot))$ is a κ-solution" if it is a κ-solution for some $\kappa > 0$.

From Appendix F, $R_t \geq 0$ for an ancient solution. This implies the essential equivalence of the notions of κ-noncollapsing in Definitions 13.14 and 26.1 , when restricted to ancient solutions. Namely, if a solution is κ-collapsed in the sense of Definition 26.1 then it is automatically κ-collapsed in the sense of Definition 13.14. Conversely, if a time-t_0 slice of an ancient solution is collapsed in the sense of Definition 13.14 then the fact that $R_t \geq 0$, together with bounds on distance distortion, implies that it is collapsed in the sense of Definition 26.1 (possibly for a different value of κ).

The relevance of κ-solutions is that a blowup limit of a finite-time singularity on a compact manifold will be a κ-solution.

For examples of κ-solutions, if $n \geq 3$ then there is a κ-solution on the cylinder $\mathbb{R} \times S^{n-1}(r)$, where the radius satisfies $r^2(t) = r_0^2 - 2(n-2)t$. There is also a κ-solution on the $\mathbb{Z}_2$-quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^{n-1}(r)$, where the generator of $\mathbb{Z}_2$ acts by reflection on $\mathbb{R}$ and by the antipodal map on S^{n-1}. On the other hand, the quotient solution on $S^1 \times S^{n-1}(r)$ is not κ-noncollapsed for any $\kappa > 0$, as can be seen by looking at large negative time.

Bryant's gradient steady soliton is a three-dimensional κ-solution given by $g(t) = \phi_t^* g_0$, where $g_0 = dr^2 + \mu(r)\, d\Theta^2$ is a certain rotationally symmetric metric on $\mathbb{R}^3$. It has sectional curvatures that go like r^{-1}, and $\mu(r) \sim r$. The gradient function f satisfies $R_{ij} + \nabla_i \nabla_j f = 0$, with $f(r) \sim -2r$. Then for r and $r - 2t$ large, $\phi_t(r, \Theta) \sim (r - 2t, \Theta)$. In particular, if $R_0 \in C^\infty(\mathbb{R}^3)$ is the scalar curvature function of g_0 then $R(t, r, \Theta) \sim R_0(r - 2t, \Theta)$.

To check the conclusion of Corollary 47.2 in this case, given a point $(r_0, \Theta) \in \mathbb{R}^3$ at time 0, the scalar curvature goes like r_0^{-1}. Multiplying the soliton metric by r_0^{-1} and sending $t \to r_0 t$ gives the asymptotic metric

$$d(r/\sqrt{r_0})^2 \; + \; \frac{r - 2r_0 t}{r_0}\, d\Theta^2. \tag{38.2}$$

Putting $u = (r - r_0)/\sqrt{r_0}$, the rescaled metric is approximately

$$du^2 \; + \; \left(1 \; + \; \frac{u}{\sqrt{r_0}} \; - \; 2t\right) d\Theta^2. \tag{38.3}$$

Given $\epsilon > 0$, this will be ϵ-biLipschitz close to the evolving cylinder $du^2 + (1 - 2t)\, d\Theta^2$ provided that $|u| \leq \epsilon\sqrt{r_0}$, i.e. $|r - r_0| \leq \epsilon r_0$. To have an ϵ-neck, we want this to hold

whenever $|r - r_0|^2 \leq (\epsilon r_0^{-1})^{-1}$. This will be the case if $r_0 \geq \epsilon^{-3}$. Thus M_ϵ is approximately

$$\{(r, \Theta) \in \mathbb{R}^3 \,:\, r \,\leq\, \epsilon^{-3}\} \tag{38.4}$$

and $Q \;=\; R(x_0, 0) \sim \epsilon^3$. Then $\mathrm{diam}(M_\epsilon) \sim \epsilon^{-3}$ and at the origin $0 \in M_\epsilon$, $R(0,0) \sim \epsilon^0$. It follows that for the value of κ corresponding to this solution, $C(\epsilon, \kappa)$ must grow at least as fast as ϵ^{-3} as $\epsilon \to 0$.

39. I.11.2. Asymptotic solitons

This section shows that every κ-solution has a gradient shrinking soliton buried inside of it, in an asymptotic sense as $t \to -\infty$. Such a soliton will be called an *asymptotic soliton.*

Heuristically, the existence of an asymptotic soliton is a consequence of the compactness results and the monotonicity of the reduced volume. Taking an appropriate sequence of spacetime points going backward in time, one constructs a limiting rescaled solution. As the limit reduced volume is constant in time, the monotonicity formula implies that this limit solution is a gradient shrinking soliton. This is the basic idea but the rigorous argument is a bit more subtle.

Pick an arbitrary point (p, t_0) in the κ-solution $(M, g(\cdot))$. Define the reduced volume $\widetilde{V}(\tau)$ and the reduced length $l(q, \tau)$ as in Section 15, by means of curves starting from (p, t_0), with $\tau = t_0 - t$. From Section 24, for each $\tau > 0$ there is some $q(\tau) \in M$ such that $l(q(\tau), \tau) \leq \frac{n}{2}$. (Note that $l \geq 0$ from the curvature assumption.)

Proposition 39.1. *(cf. Proposition I.11.2) There is a sequence $\overline{\tau}_i \to \infty$ so that if we consider the solution $g(\cdot)$ on the time interval $[t_0 - \overline{\tau}_i, t_0 - \frac{1}{2}\overline{\tau}_i]$ and parabolically rescale it at the point $(q(\overline{\tau}_i), t_0 - \overline{\tau}_i)$ by the factor $\overline{\tau}_i^{-1}$ then as $i \to \infty$, the rescaled solutions converge to a nonflat gradient shrinking soliton (restricted to $[-1, -\frac{1}{2}]$).*

Proof. Equation (25.5) implies that $|\nabla l^{1/2}|^2 \,\leq\, \frac{C}{4\tau}$, and so

$$|l^{1/2}(q, \tau) \,-\, l^{1/2}(q(\tau), \tau)| \,\leq\, \sqrt{\frac{C}{4\tau}}\ \mathrm{dist}_{t_0 - \tau}(q, q(\tau)). \tag{39.2}$$

We apply this estimate initially at some fixed time $\tau = \overline{\tau}$, to obtain

$$l(q, \overline{\tau}) \,\leq\, \left(\sqrt{\frac{C}{4\overline{\tau}}}\ \mathrm{dist}_{t_0 - \overline{\tau}}(q, q(\overline{\tau})) \,+\, \sqrt{\frac{n}{2}} \right)^2. \tag{39.3}$$

From (18.13), (18.14) and (25.5),

$$\partial_\tau l \,=\, \frac{R}{2} \,-\, \frac{|\nabla l|^2}{2} \,-\, \frac{l}{2\tau} \,\geq\, -\, \frac{(1+C)l}{2\tau}. \tag{39.4}$$

This implies that for $\tau \in \left[\frac{1}{2}\overline{\tau}, \overline{\tau}\right]$,

$$l(q, \tau) \,\leq\, \left(\frac{\overline{\tau}}{\tau}\right)^{\frac{1+C}{2}} \left(\sqrt{\frac{C}{4\overline{\tau}}}\ \mathrm{dist}_{t_0 - \overline{\tau}}(q, q(\overline{\tau})) \,+\, \sqrt{\frac{n}{2}} \right)^2. \tag{39.5}$$

Also from (25.5), we have $\tau R \,\leq\, Cl$. Then we can plug in the previous bound on l to get an upper bound on τR for $\tau \in \left[\frac{1}{2}\overline{\tau}, \overline{\tau}\right]$. The upshot is that for any $\epsilon > 0$, one can find

$\delta > 0$ so that both $l(q,\tau)$ and $\tau R(q, t_0 - \tau)$ do not exceed δ^{-1} whenever $\frac{1}{2}\overline{\tau} \leq \tau \leq \overline{\tau}$ and $\mathrm{dist}^2_{t_0-\overline{\tau}}(q, q(\overline{\tau})) \leq \epsilon^{-1}\overline{\tau}$.

Varying $\overline{\tau}$, as the rescaled solutions (with basepoints at $(q(\overline{\tau}), t_0 - \overline{\tau})$) are uniformly noncollapsing and have uniform curvature bounds on balls, Appendix E implies that we can take a sequence $\overline{\tau}_i \to \infty$ to get a pointed limit $(\overline{M}, \overline{q}, \overline{g}(\cdot))$ that is a complete Ricci flow solution (in the backward parameter τ) for $\frac{1}{2} < \tau < 1$. We may assume that we have locally Lipschitz convergence of l to a limit function $\overline{l}$.

We define the reduced volume $\tilde{V}(\tau)$ for the limit solution using the limit function $\overline{l}$. We claim that for any $\tau \in \left(\frac{1}{2}, 1\right)$, if we put $\tau_i = \tau\overline{\tau}_i$ then the number $\tilde{V}(\tau)$ for the limit solution is the limit of numbers $\tilde{V}(\tau_i)$ for the original solution. One wishes to apply dominated convergence to the integrals $\int_M e^{-l(q,\tau_i)}\tau_i^{-n/2}\,\mathrm{dvol}(q, t_0 - \tau_i)$. (Note that $\tau_i^{-n/2}\,\mathrm{dvol}(q, t_0 - \tau_i) = \tau^{-n/2}\overline{\tau}_i^{-n/2}\,\mathrm{dvol}(q, t_0 - \tau_i)$ and $\overline{\tau}_i^{-n/2}\,\mathrm{dvol}(q, t_0 - \tau_i)$ is the volume form for the rescaled metric $\overline{\tau}_i^{-1}\, g(t_0 - \tau_i)$.) However, to do so one needs uniform lower bounds on $l(q, \tau')$ for the original solution in terms of $d_{t_0-\tau'}(q, q(\tau'))$, for $\tau' \in (-\infty, 0)$. By an argument of Perelman, written in detail in [67], one does indeed have a lower bound of the form

$$l(q,\tau') \geq -l(q(\tau')) - 1 + C(n)\,\frac{d_{t_0-\tau'}(q, q(\tau'))^2}{\tau'}. \tag{39.6}$$

The nonnegative curvature gives polynomial volume growth for distance balls, so using (39.6) one can apply dominated convergence to the integrals $\int_M e^{-l(q,\tau_i)}\,\tau_i^{-n/2}\,\mathrm{dvol}(q, t_0 - \tau_i)$. Thus $\lim_{i\to\infty} \tilde{V}(\tau_i) = \tilde{V}(\tau)$.

As (39.5) gives a uniform upper bound on l on an appropriate ball around $q(\tau_i)$, and there is a lower volume bound on the ball, it follows that as $i \to \infty$, $\tilde{V}(\tau_i)$ is uniformly bounded away from zero. From this argument and the monotonicity of $\tilde{V}$, $\tilde{V}(\tau)$ is a positive constant c as a function of τ, namely the limit of the reduced volume of the original solution as real time goes to $-\infty$. As the original solution is nonflat, the constant c is strictly less than the limit of the reduced volume of the original solution as real time goes to zero, which is $(4\pi)^{\frac{n}{2}}$.

Next, we will apply (24.6) and (24.8). As (24.6) holds distributionally for each rescaled solution, it follows that it holds distributionally for $\overline{l}$. In particular, the nonpositivity implies that the left-hand side of (24.6), when computed for the limit solution, is actually a nonpositive measure. If the left-hand side of (24.6) (for the limit solution) were not strictly zero then using (24.7) we would conclude that $\frac{d\tilde{V}}{d\tau}$ is somewhere negative, which is a contradiction. (We use the fact that (39.6) passes to the limit to give a similar lower bound on $\overline{l}$.) Thus we must have equality in (24.6) for the limit solution. This implies equality in (21.2), which implies equality in (24.8). Writing (24.8) as

$$(4\triangle - R)\, e^{-\frac{l}{2}} = \frac{l - n}{\tau}\, e^{-\frac{l}{2}}, \tag{39.7}$$

elliptic theory gives smoothness of $\overline{l}$.

In I.11.2 it is said that equality in (24.8) implies equality in (23.9), which implies that one has a gradient shrinking soliton. There is a problem with this argument, as the use of (23.9) implicitly assumes that the solution is defined for all $\tau \geq 0$, which we do not know. (The function $\overline{l}$ is only defined by a limiting procedure, and not in terms of $\mathcal{L}$-geodesics on some

Ricci flow solution.) However, one can instead use Proposition 29.5, with $f = \overline{l}$. Equality in (24.8) implies that $v = 0$, so (29.6) directly gives the gradient shrinking soliton equation. (The problem with the argument using (23.9), and its resolution using (29.6), were pointed out by the UCSB group.)

If the gradient shrinking soliton $\overline{g}(\cdot)$ is flat then, as it will be κ-noncollapsed at all scales, it must be $\mathbb{R}^n$. From the soliton equation, $\partial_i \partial_j \overline{l} = \frac{\overline{g}_{ij}}{2\tau}$ and $\triangle \overline{l} = \frac{n}{2\tau}$. Putting this into the equality (24.8) gives $|\nabla \overline{l}|^2 = \frac{\overline{l}}{\tau}$. It follows that the level sets of $\overline{l}$ are distance spheres. Then (24.6) implies that with an appropriate choice of origin, $\overline{l} = \frac{|x|^2}{4\tau}$. The reduced volume $\tilde{V}(\tau)$ for the limit solution is now computed to be $(4\pi)^{\frac{n}{2}}$, which is a contradiction. Therefore the gradient shrinking soliton is not flat. □

We remark that the gradient soliton constructed here does not, *a priori*, have bounded curvature on compact time intervals, i.e. it may not be a κ-solution. In the 2 and 3-dimensional cases one can prove this using additional reasoning. See Section 43 where it is shown that 2-dimensional κ-solutions are round spheres, and Section 46 where the 3-dimensional case is discussed.

40. I.11.3. Two dimensional κ-solutions

The next result is a classification of two-dimensional κ-solutions. It is important when doing dimensional reduction.

Corollary 40.1. *(cf. Corollary I.11.3) The only oriented two-dimensional κ-solution is the shrinking round 2-sphere.*

Proof. First, the only nonflat oriented nonnegatively curved gradient shrinking 2-D soliton is the round S^2. The reference [30] given in I.11.3 for this fact does not actually cover it, as the reference only deals with compact solitons. A proof using Proposition 39.1 to rule out the noncompact case appears in [68].

Given this, the limit solution in Proposition 39.1 is a shrinking round 2-sphere. Thus the rescalings $\overline{\tau}_i^{-1} g(t_0 - \overline{\tau}_i)$ converge to a round 2-sphere as $i \to \infty$. However, by [30] the Ricci flow makes an almost-round 2-sphere become more round. Thus any given time slice of the original κ-solution must be a round 2-sphere. □

Remark 40.2. One can employ a somewhat different line of reasoning to prove Corollary 40.1; see Section 43.

41. I.11.4. Asymptotic scalar curvature and asymptotic volume ratio

In this section we first show that the asymptotic scalar curvature ratio of a κ-solution is infinite. We then show that the asymptotic volume ratio vanishes. The proofs are somewhat rearranged from those in I.11.4. They are logically independent of Section 40, i.e. also cover the case $n = 2$. We will use results from Appendices F and G, in particular (F.14).

Definition 41.1. If M is a complete connected Riemannian manifold then its *asymptotic scalar curvature ratio* is $\mathcal{R} = \limsup_{x\to\infty} R(x)\, d(x,p)^2$. It is independent of the choice of basepoint p.

Theorem 41.2. *Let $(M, g(\cdot))$ be a noncompact κ-solution. Then the asymptotic scalar curvature ratio $\mathcal{R}$ is infinite for each time slice.*

Proof. Suppose that M is n-dimensional, with $n \geq 2$. Pick $p \in M$ and consider a time-t_0 slice $(M, g(t_0))$. We deal with the cases $\mathcal{R} \in (0, \infty)$ and $\mathcal{R} = 0$ separately and show that they lead to contradictions.

Case 1: $0 < \mathcal{R} < \infty$. We choose a sequence $x_k \in M$ such that $d_{t_0}(x_k, p) \to \infty$ and $R(x_k, t_0)d^2(x_k, p) \to \mathcal{R}$. Consider the rescaled pointed solution $(M, x_k, g_k(t))$ with $g_k(t) = R(x_k, t_0)g(t_0 + \frac{t}{R(x_k,t_0)})$ and $t \in (-\infty, 0]$. We have $R_k(x_k, 0) = 1$, and for all $b > 0$, for sufficiently large k, we have $R_k(x, t) \leq R_k(x, 0) \leq \frac{2\mathcal{R}}{d_k^2(x,p)}$ for all x such that $d_k(x, p) > b$. Fix numbers $b, B > 0$ so that $b < \sqrt{\mathcal{R}} < B$. The κ-noncollapsing assumption gives a uniform positive lower bound on the injectivity radius of $g_k(0)$ at x_k, and so by Appendix E we may extract a pointed limit solution $(M_\infty, x_\infty, g_\infty(\cdot))$, defined on a time interval $(-\infty, 0]$ from the sequence $(M_k, x_k, g_k(\cdot))$ where $M_k = \{x \in M \mid b < d_k(x, p) < B\}$. Note that g_∞ has nonnegative curvature operator and the time slice $(M_\infty, g_\infty(0))$ is locally isometric to an annular portion of a nonflat metric cone, since $(M_k, p, g_k(0))$ Gromov-Hausdorff converges to the Tits cone $C_T(M, g(t_0))$. (We use the word "locally" because the annulus in $C_T(M, g(t_0))$ need not be geodesically convex in $C_T(M, g(t_0))$, so we are only saying that the distance functions in small balls match up.) When $n = 2$ this contradicts the fact that $R_\infty(x_\infty, 0) = 1$. When $n \geq 3$, we will derive a contradiction from Hamilton's curvature evolution equation

$$\mathrm{Rm}_t = \Delta\,\mathrm{Rm} + Q(\mathrm{Rm}). \tag{41.3}$$

Let $d_v : C_T(M, g(t_0)) \to \mathbb{R}$ be the distance function from the vertex and let $\rho : M_\infty \to \mathbb{R}$ be the pullback of d_v under the inclusion of the annulus M_∞ in $C_T(M, g(t_0))$.

Lemma 41.4. *The metric cone structure on $(M_\infty, g_\infty(0))$ is smooth, i.e. ρ is a smooth function.*

Proof. Consider a unit speed geodesic segment γ in the Tits cone $C_T(M, g(t_0))$, such that γ is disjoint from the vertex $v \in C_T(M, g(t_0))$. Note that since $C_T(M, g(t_0))$ is a Euclidean cone over the Tits boundary $\partial_T(M, g(t_0))$, the geodesic γ lies in the cone over a geodesic segment $\hat{\gamma} \subset \partial_T(M, g(t_0))$. Thus γ lies in a 2-dimensional locally convex flat subspace of $C_T(M, g(t_0))$. Also, as in a flat 2-dimensional cone, the second derivative of the composite function $d_v^2 \circ \gamma$ is identically 2.

Since ρ is obtained from d_v by composition with a locally isometric embedding $(M_\infty, g_\infty(0)) \to C_T(M, g(t_0))$, the composition of ρ^2 with any unit speed geodesic segment in M_∞ also has second derivative identically equal to 2.

Since ρ^2 is Lipschitz, Rademacher's theorem implies that ρ^2 is differentiable almost everywhere. Let $y \in M_\infty$ be a point of differentiability of ρ^2. If the injectivity radius of $g_\infty(0)$

at y is $> a$, then the function h_y given by the composition

$$T_yM_\infty \supset B(0,a) \xrightarrow{\exp_y} M_\infty \xrightarrow{\rho^2} \mathbb{R}$$

has the property that its second radial derivative is identically 2, and it is differentiable at the origin $0 \in T_yM_\infty$. Therefore h_y is a second order polynomial with Hessian identically 2, and is smooth. As the injectivity radius is a continuous function, this implies that ρ^2 is smooth everywhere in M_∞.

Since ρ^2 is strictly positive on M_∞, it follows that $\rho = \sqrt{\rho^2}$ is smooth as well. $\square$

By the lemma, we may choose a smooth local orthonormal frame $e_1, \ldots, e_n$ for $(M_\infty, g_\infty(0))$ near x_∞ such that e_1 points radially outward (with respect to the cone structure), and e_2, e_3 span a 2-plane at x_∞ with strictly positive curvature; such a 2-plane exists because $R_\infty(x_\infty, 1) = 1$. Put $P = e_1 \wedge e_2$. In terms of the curvature operator, the fact that $\mathrm{Rm}_\infty(e_1, e_2, e_2, e_1) = 0$ is equivalent to $\langle P, \mathrm{Rm}_\infty P\rangle = 0$. As the curvature operator is nonnegative, it follows that $\mathrm{Rm}_\infty P = 0$. (In fact, this is true for any metric cone.) Differentiating gives

$$(\nabla_{e_i} \mathrm{Rm}_\infty)\, P \,+\, \mathrm{Rm}_\infty(\nabla_{e_i} P) \,=\, 0 \tag{41.5}$$

and

$$(\triangle\, \mathrm{Rm}_\infty)\, P \,+\, 2 \sum_i (\nabla_{e_i} \mathrm{Rm}_\infty)\, \nabla_{e_i} P \,+\, \mathrm{Rm}_\infty\, (\triangle P) \,=\, 0. \tag{41.6}$$

Taking the inner product of (41.6) with P gives

$$\begin{aligned} 0 \,&=\, \langle P, (\triangle\, \mathrm{Rm}_\infty)\, P\rangle \,+\, 2 \sum_i \langle P, (\nabla_{e_i} \mathrm{Rm}_\infty)\, \nabla_{e_i} P\rangle \\ &=\, \langle P, (\triangle\, \mathrm{Rm}_\infty)\, P\rangle \,+\, 2 \sum_i \langle \nabla_{e_i} P, (\nabla_{e_i} \mathrm{Rm}_\infty)\, P\rangle. \end{aligned} \tag{41.7}$$

Then (41.5) gives

$$\langle P, (\triangle\, \mathrm{Rm}_\infty) P\rangle \,=\, 2 \sum_i \langle \nabla_{e_i} P, \mathrm{Rm}_\infty\, (\nabla_{e_i} P)\rangle. \tag{41.8}$$

As the sphere of distance r from the vertex in a metric cone has principal curvatures $\frac{1}{r}$, we have $\nabla_{e_3} e_1 \,=\, -\,\frac{1}{r}\, e_3$. Then

$$\nabla_{e_3}(e_1 \wedge e_2) \,=\, (\nabla_{e_3} e_1) \wedge e_2 \,+\, e_1 \wedge \nabla_{e_3} e_2 \,=\, \frac{1}{r}\, (e_2 \wedge e_3) \,+\, e_1 \wedge \nabla_{e_3} e_2. \tag{41.9}$$

This shows that $\nabla_{e_3} P$ has a nonradial component $\frac{1}{r} e_2 \wedge e_3$. Thus $(\Delta\, \mathrm{Rm}_\infty)(e_1, e_2, e_2, e_1) > 0$. The zeroth order quadratic term $Q(\mathrm{Rm})$ appearing in (41.3) is nonnegative when Rm is nonnegative, so we conclude that $\partial_t \mathrm{Rm}_\infty(e_1, e_2, e_2, e_1) > 0$ at $t = 0$. This means that $\mathrm{Rm}_\infty(-\epsilon)(e_1, e_2, e_2, e_1) < 0$ for $\epsilon > 0$ sufficiently small, which is impossible.

Case 2: $\mathcal{R} = 0$. Let us take any sequence $x_k \in M$ with $d_{t_0}(x_k, p) \to \infty$. Set $r_k = d_{t_0}(x_k, p)$, put

$$g_k(t) = r_k^{-2} g(t_0 + r_k^2 t) \tag{41.10}$$

for $t \in (-\infty, 0]$, and let $d_k(\cdot, \cdot)$ be the distance function associated to $g_k(0)$. For any $0 < b < B$, put

$$M_k(b, B) = \{x \in M \mid 0 < b < d_k(x, p) < B\}. \tag{41.11}$$

Since $\mathcal{R} = 0$, we get that $\sup_{x \in M_k(b,B)} |\operatorname{Rm}_k(x, 0)| \to 0$ as $k \to \infty$. Invoking the κ-noncollapsed assumption as in the previous case, we may assume that $(M, p, g_k(0))$ Gromov-Hausdorff converges to a metric cone $(M_\infty, p_\infty, g_\infty)$ (the Tits cone $C_T(M, g(t_0))$) which is flat and smooth away from the vertex p_∞, and the convergence is smooth away from p_∞.

The "unit sphere" in $C_T(M, g(t_0))$ defines a compact smooth hypersurface S_∞ in $(M_\infty - \{p_\infty\}, g_\infty(0))$ whose principal curvatures are identically 1. If $n \geq 3$ then S_∞ must be a quotient of the standard $(n-1)$-sphere by the free action of a finite group of isometries. We have a sequence $S_k \subset M_k$ of approximating smooth hypersurfaces whose principal curvatures (with respect to $g_k(0)$) go to 1 as $k \to \infty$. In view of the convergence to $(M_\infty, p_\infty, g_\infty)$, for sufficiently large k, the inward principal curvatures of S_k with respect to $g_k(0)$ are close to 1. As M has nonnegative curvature, S_k is diffeomorphic to a sphere [28, Theorem A]. Thus S_∞ is isometric to the standard $(n-1)$-sphere, and so $C_T(M, g(t_0))$ is isometric to n-dimensional Euclidean space. Then $(M, g(t_0))$ is isometric to $\mathbb{R}^n$, which contradicts the definition of a κ-solution.

In the case $n = 2$ we know that S_∞ is diffeomorphic to a circle but we do not know *a priori* that it has length 2π. To handle the case $n = 2$, we use the fact that $g_k(t)$ is a Ricci flow solution, to extract a limiting smooth incomplete time-independent Ricci flow solution $(M_\infty \backslash p_\infty, g_\infty(t))$ for $t \in [-1, 0]$. Note that this solution is unpointed. In view of the convergence to the limiting solution, for sufficiently large k, the inward principal curvatures of S_k with respect to $g_k(t)$ are close to 1 for all $t \in [-1, 0]$. This implies that S_k bounds a domain $B_k \subset M$ whose diameter with respect to $g_k(t)$ is uniformly bounded above, say by 10 (see Appendix G).

Applying the Harnack inequality (F.14) with $y_k \in S_k$ (at time 0) and $x \in B_k$ (at time -1), we see that $\sup_{x \in B_k} |\operatorname{Rm}_k(x, -1)| \to 0$ as $k \to \infty$. Thus $(B_k, p, g_k(-1))$ Gromov-Hausdorff converges to a flat manifold $(B_\infty, \bar{p}_\infty, g_\infty(-1))$ with convex boundary. As all of the principal curvatures of ∂B_∞ are 1, B_∞ must be isometric to a Euclidean unit ball. This implies that S_∞ is isometric to the standard S^1 of length 2π, and we obtain a contradiction as before. $\square$

Definition 41.12. If M is a complete n-dimensional Riemannian manifold with nonnegative Ricci curvature then its *asymptotic volume ratio* is $\mathcal{V} = \lim_{r \to \infty} r^{-n} \operatorname{vol}(B(p, r))$. It is independent of the choice of basepoint x_0.

Proposition 41.13. *(cf. Proposition I.11.4) Let $(M, g(\cdot))$ be a noncompact κ-solution. Then the asymptotic volume ratio $\mathcal{V}$ vanishes for each time slice $(M, g(t_0))$. Moreover, there is a sequence of points $x_k \in M$ going to infinity such that the pointed sequence $\{(M, (x_k, t_0), g(\cdot))\}_{k=1}^\infty$ converges, modulo rescaling by $R(x_k, t_0)$, to a κ-solution which isometrically splits off an $\mathbb{R}$-factor.*

Proof. Consider the time-t_0 slice. Suppose that $\mathcal{V} > 0$. As $\mathcal{R} = \infty$, there are sequences $x_k \in M$ and $s_k > 0$ such that $d_{t_0}(x_k, p) \to \infty$, $\frac{s_k}{d_{t_0}(x_k, p)} \to 0$, $R(x_k, t_0) s_k^2 \to \infty$, and $R(x, t_0) \leq 2R(x_k, t_0)$ for all $x \in B_{t_0}(x_k, s_k)$ [33, Lemma 22.2]. Consider the rescaled pointed

solution $(M, x_k, g_k(t))$ with $g_k(t) = R(x_k, t_0)\, g(t_0 + \frac{t}{R(x_k,t_0)})$ and $t \in (-\infty, 0]$. As $R_t \geq 0$, we have $R_k(x,t) \leq 2$ whenever $t \leq 0$ and $d_k(x, x_k) \leq R(x_k, t_0)^{1/2} s_k$, where d_k is the distance function for $g_k(0)$ and $R_k(\cdot, \cdot)$ is the scalar curvature of $g_k(\cdot)$. The κ-noncollapsing assumption gives a uniform positive lower bound on the injectivity radius of $g_k(0)$ at x_k, so by Appendix E we may extract a complete pointed limit solution $(M_\infty, x_\infty, g_\infty(t))$, $t \in (-\infty, 0]$, of a subsequence of the sequence of pointed Ricci flows. By relative volume comparison, $(M_\infty, x_\infty, g_\infty(0))$ has positive asymptotic volume ratio. By Appendix G, the Riemannian manifold $(M_\infty, x_\infty, g_\infty(0))$ is isometric to an Alexandrov space which splits off a line, which means that it is a Riemannian product $\mathbb{R} \times N$. This implies a product structure for earlier times; see Appendix A. Now when $n = 2$, we have a contradiction, since $R(x_\infty, 0) = 1$ but $(M_\infty, g_\infty(0))$ is a product surface, and must therefore be flat. When $n > 2$ we obtain a κ-solution on an $(n-1)$-manifold with positive asymptotic volume ratio at time zero, and by induction this is impossible. □

42. In a κ-solution, the curvature and the normalized volume control each other

In this section we show that, roughly speaking, in a κ-solution the curvature and the normalized volume control each other.

Corollary 42.1. *1. If $B(x_0, r_0)$ is a ball in a time slice of a κ-solution, then the normalized volume $r_0^{-n} \operatorname{vol}(B(x_0, r_0))$ is controlled (i.e. bounded away from zero) $\Longleftrightarrow$ the normalized scalar curvature $r_0^2 R(x_0)$ is controlled (i.e. bounded above).*

2. If $B(x_0, r_0)$ is a ball in a time slice of a κ-solution, then the normalized volume $r_0^{-n} \operatorname{vol}(B(x_0, r_0))$ is almost maximal $\Longleftrightarrow$ the normalized scalar curvature $r_0^2 R(x_0)$ is almost zero.

3. (Precompactness) If $(M_k, (x_k, t_k), g_k(\cdot))$ is a sequence of pointed κ-solutions (without the assumption that $R(x_k, t_k) = 1$) and for some $r > 0$, the r-balls $B(x_k, r) \subset (M_k, g_k(t_k))$ have controlled normalized volume, then a subsequence converges to an ancient solution $(M_\infty, (x_\infty, 0), g_\infty(\cdot))$ which has nonnegative curvature operator, and is κ-noncollapsed (though a priori the curvature may be unbounded on a given time slice).

4. There is a constant $\eta = \eta(n, \kappa)$ such that for every n-dimensional κ-solution $(M, g(\cdot))$, and all $x \in M$, we have $|\nabla R|(x,t) \leq \eta R^{\frac{3}{2}}(x,t)$ and $|R_t|(x,t) \leq \eta R^2(x,t)$. More generally, there are scale invariant bounds on all derivatives of the curvature tensor, that only depend on n and κ. That is, for each $\rho, k, l < \infty$ there is a constant $C = C(n, \rho, k, l, \kappa) < \infty$ such that $\left| \frac{\partial^k}{\partial t^k} \nabla^l \operatorname{Rm} \right| (y,t) \leq C\, R(x,t)^{(k + \frac{l}{2} + 1)}$ for any $y \in B_t(x, \rho R(x,t)^{-\frac{1}{2}})$.

5. There is a function $\alpha : [0, \infty) \to [0, \infty)$ depending only on κ such that $\lim_{s \to \infty} \alpha(s) = \infty$, and for every κ-solution $(M, g(\cdot))$ and $x, y \in M$, we have $R(y) d^2(x,y) \geq \alpha(R(x) d^2(x,y))$.

Proof. Assertion 1, $\Longrightarrow$. Suppose we have a sequence of κ-solutions $(M_k, g_k(\cdot))$, and sequences $t_k \in (-\infty, 0]$, $x_k \in M_k$, $r_k > 0$, such that at time t_k, the normalized volume of $B(x_k, r_k)$ is $\geq c > 0$, and $R(x_k, t_k) r_k^2 \to \infty$. By Appendix H, for each k, we can find $y_k \in B(x_k, 5r_k)$, $\bar{r}_k \leq r_k$, such that $R(y_k, t_k) \bar{r}_k^2 \geq R(x_k, t_k) r_k^2$, and $R(z, t_k) \leq 2R(y_k, t_k)$ for

all $z \in B(y_k, \bar{r}_k)$. Note that by relative volume comparison, whenever $\widetilde{r}_k \leq \bar{r}_k$ we have

$$\frac{\operatorname{vol}(B(y_k, \widetilde{r}_k))}{\widetilde{r}_k^n} \geq \frac{\operatorname{vol}(B(y_k, \bar{r}_k))}{\bar{r}_k^n} \geq \frac{\operatorname{vol}(B(y_k, 10r_k))}{(10r_k)^n} \geq \frac{\operatorname{vol}(B(x_k, r_k))}{(10r_k)^n} \geq \frac{c}{10^n}. \tag{42.2}$$

Rescaling the sequence of pointed solutions $(M_k, (y_k, t_k), g_k(\cdot))$ by $R(y_k, t_k)$, we get a sequence satisfying the hypotheses of Appendix E (we use here the fact that $R_t \geq 0$ for an ancient solution), so it accumulates on a limit flow $(M_\infty, (y_\infty, 0), g_\infty(\cdot))$ which is a κ-solution. By (42.2), the asymptotic volume ratio of $(M, g_\infty(0))$ is $\geq \frac{c}{10^n} > 0$. This contradicts Proposition 41.13.

Assertion 3. By relative volume comparison, it follows that every r-ball in $(M_k, g_k(t_k))$ has normalized volume bounded below by a (k-independent) function of its distance to x_k. By 1, this implies that the curvature of $(M_k, g_k(t_k))$ is bounded by a k-independent function of the distance to x_k, and hence we can apply Appendix E to extract a smoothly converging subsequence.

Assertion 1, $\Longleftarrow$. Suppose we have a sequence $(M_k, g_k(\cdot))$ of κ-solutions, and sequences $x_k \in M_k$, $r_k > 0$, such that $R(x_k, t_k) r_k^2 < c$ for all k, but $r_k^{-n} \operatorname{vol}(B(x_k, r_k)) \to 0$. For large k, we can choose $\bar{r}_k \in (0, r_k)$ such that $\bar{r}_k^{-n} \operatorname{vol}(B(x_k, \bar{r}_k)) = \frac{1}{2} c_n$ where c_n is the volume of the unit Euclidean n-ball. By relative volume comparison, $\frac{\bar{r}_k}{r_k} \to 0$. Applying 3, we see that the pointed sequence $(M_k, (x_k, t_k), g_k(\cdot))$, rescaled by the factor $\bar{r}_k^{-2}$, accumulates on a pointed ancient solution $(M_\infty, (x_\infty, 0), g_\infty(\cdot))$, such that the ball $B(x_\infty, 1) \subset (M_\infty, g_\infty)$ has normalized volume $\frac{1}{2} c_n$ at $t = 0$.

Suppose the ball $B(x_\infty, 1) \subset (M_\infty, g_\infty(0))$ were flat. Then by the Harnack inequality (F.14) (applied to the approximators) we would have $R_\infty(x, t) = 0$ for all $x \in M_\infty$, $t \leq 0$, i.e. $(M_\infty, g_\infty(t))$ would be a time-independent flat manifold. It cannot be $\mathbb{R}^n$ since $\operatorname{vol}(B(x_\infty, 1)) = \frac{1}{2} c_n$. But flat manifolds other than Euclidean space have zero asymptotic volume ratio (as follows from the Bieberbach theorem that if $N = \mathbb{R}^n/\Gamma$ is a flat manifold and Γ is nontrivial then there is a Γ-invariant affine subspace $A \subset \mathbb{R}^n$ of dimension at least 1 on which Γ acts cocompactly). This contradicts the assumption that the sequence $(M_k, g_k(\cdot))$ is κ-noncollapsed. Thus $B(x_\infty, 1) \subset (M_\infty, g_\infty(0))$ is not flat, which means, by the Harnack inequality, that the scalar curvature of $g_\infty(0)$ is strictly positive everywhere. Therefore, with respect to g_k, we have

$$\liminf_{k \to \infty} R(x_k, t_k) r_k^2 = \liminf_{k \to \infty} (R(x_k, t_k)) \bar{r}_k^2) \left(\frac{r_k}{\bar{r}_k} \right)^2 \geq \text{const.} \liminf_{k \to \infty} \left(\frac{r_k}{\bar{r}_k} \right)^2 = \infty, \tag{42.3}$$

which is a contradiction.

Assertion 2, $\Longrightarrow$. Apply 1, the precompactness criterion, and the fact that a nonnegatively-curved manifold whose balls have normalized volume c_n must be flat.

Assertion 2, $\Longleftarrow$. Apply 1, the precompactness criterion, and the Harnack inequality (F.14) (to the approximators).

Assertion 4. This follows by rescaling g so that $R(x, t) = 1$, and applying 1 and 3.

Assertion 5. The quantity $R(z)d^2(u,v)$ is scale invariant. If the assertion failed then we would have sequences $(M_k, g_k(\cdot))$, $x_k, y_k \in M_k$, such that $R(y_k) = 1$ and $d(x_k, y_k)$ remains bounded, but the curvature at x_k blows up. This contradicts 1 and 3.

43. An alternate proof of Corollary 40.1 using Proposition 41.13 and Corollary 42.1

In this section we give an alternate proof of Corollary 40.1. It uses Proposition 41.13 and Corollary 42.1 To clarify the chain of logical dependence, we remark that this section is concerned with 2-dimensional κ-solutions, and does not use anything from Sections 39 or 40. It does use Proposition 41.13. However, we avoid circularity here because the proof of Proposition 41.13 given in Section 41, unlike the proof in [51], does not use Corollary 40.1.

Lemma 43.1. *There is a constant $v = v(\kappa) > 0$ such that if $(M, g(\cdot))$ is a 2-dimensional κ-solution (a priori either compact or noncompact), $x, y \in M$ and $r = d(x, y)$ then*

$$\mathrm{vol}(B_t(x, r)) \geq vr^2. \tag{43.2}$$

Proof. If the lemma were not true then there would be a sequence $(M_k, g_k(\cdot))$ of 2-dimensional κ-solutions, and sequences $x_k, y_k \in M_k$, $t_k \in \mathbb{R}$ such that $r_k^{-2}\,\mathrm{vol}(B_{t_k}(x_k, r_k)) \to 0$, where $r_k = d(x_k, y_k)$. Let z_k be the midpoint of a shortest segment from x_k to y_k in the t_k-time slice $(M_k, g_k(t_k))$. For large k, choose $\bar{r}_k \in (0, r_k/2)$ such that

$$\bar{r}_k^{-2}\,\mathrm{vol}(B_{t_k}(z_k, \bar{r}_k)) = \frac{\pi}{2}, \tag{43.3}$$

i.e. half the area of the unit disk in $\mathbb{R}^2$. As

$$\frac{\pi}{2} = \bar{r}_k^{-2}\,\mathrm{vol}(B_{t_k}(z_k, \bar{r}_k)) \leq \bar{r}_k^{-2}\,\mathrm{vol}(B_{t_k}(x_k, r_k)) = (\bar{r}_k/r_k)^{-2}\, r_k^{-2}\,\mathrm{vol}(B_{t_k}(x_k, r_k)), \tag{43.4}$$

it follows that $\lim_{k\to\infty} \frac{\bar{r}_k}{r_k} = 0$. Then by part 3 of Corollary 42.1, the sequence of pointed Ricci flows $(M_k, (z_k, t_k), g_k(\cdot))$, when rescaled by $\bar{r}_k^{-2}$, accumulates on a complete Ricci flow $(M_\infty, (z_\infty, 0), g_\infty(\cdot))$. The segments from z_k to x_k and y_k accumulate on a line in $(M_\infty, g_\infty(0))$, and hence $(M_\infty, g_\infty(0))$ splits off a line. By (43.3), $(M_\infty, g_\infty(0))$ cannot be isometric to $\mathbb{R}^2$, and hence must be a cylinder. Considering the approximating Ricci flows, we get a contradiction to the κ-noncollapsing assumption. □

Lemma 43.1 implies that the asymptotic volume ratio of any noncompact 2-dimensional κ-solution is at least $v > 0$. By Proposition 41.13 we therefore conclude that every 2-dimensional κ-solution is compact. (This was implicitly assumed in the proof of Corollary I.11.3 in [51], as its reference [30] is about compact surfaces.)

Consider the family $\mathcal{F}$ of 2-dimensional κ-solutions $(M, (x, 0), g(\cdot))$ with $\mathrm{diam}(M, g(0)) = 1$. By Lemma 43.1, there is uniform lower bound on the volume of the $t = 0$ time slices of κ-solutions in $\mathcal{F}$. Thus $\mathcal{F}$ is compact in the smooth topology by part 3 of Corollary 42.1 (the precompactness leads to compactness in view of the diameter bound). This implies (recall that $R > 0$) that there is a constant $K \geq 1$ such that every time slice of every 2-dimensional κ-solution has K-pinched curvature.

Hamilton has shown that volume-normalized Ricci flow on compact surfaces with positively pinched initial data converges exponentially fast to a constant curvature metric [30]. His argument shows that there is a small $\epsilon > 0$, depending continuously on the initial data, so that when the volume of the (unnormalized) solution has gone down by a factor of at least ϵ^{-1}, the pinching is at most the square root of the initial pinching. By the compactness of the family $\mathcal{F}$, this ϵ can be chosen uniformly when we take the initial data to be the $t = 0$ time slice of a κ-solution in $\mathcal{F}$.

Now let K_0 be the worst pinching of a 2-dimensional κ-solution, and let $(M, g(\cdot))$ be a κ-solution where the curvature pinching of $(M, g(0))$ is K_0. Choosing $t < 0$ such that $\epsilon \operatorname{vol}(M, g(t)) = \operatorname{vol}(M, g(0))$, the previous paragraph implies the curvature pinching of $(M, g(t))$ is at least K_0^2. This would contradict the fact that K_0 is the upper bound on the pinching for all κ-solutions, unless $K_0 = 1$.

44. I.11.5. A volume bound

In this section we give a consequence of Proposition 41.13 concerning the volumes of metric balls in Ricci flow solutions with nonnegative curvature operator.

Corollary 44.1. *(cf. Corollary I.11.5) For every $\epsilon > 0$, there is an $A < \infty$ with the following property. Suppose that we have a sequence of (not necessarily complete) Ricci flow solutions $g_k(\cdot)$ with nonnegative curvature operator, defined on $M_k \times [t_k, 0]$, such that*
1. For each k, the time-zero ball $B(x_k, r_k)$ has compact closure in M_k.
2. For all $(x,t) \in B(x_k, r_k) \times [t_k, 0]$, $\frac{1}{2}R(x,t) \le R(x_k, 0) = Q_k$.
3. $\lim_{k\to\infty} t_k Q_k = -\infty$.
4. $\lim_{k\to\infty} r_k^2 Q_k = \infty$.
Then for large k, $\operatorname{vol}(B(x_k, AQ_k^{-\frac{1}{2}})) \le \epsilon (AQ_k^{-\frac{1}{2}})^n$ at time zero.

Proof. Given $\epsilon > 0$, suppose that the corollary is not true. Then there is a sequence of such Ricci flow solutions with $\operatorname{vol}(B(x_k, A_k Q_k^{-\frac{1}{2}})) > \epsilon (A_k Q_k^{-\frac{1}{2}})^n$ at time zero, where $A_k \to \infty$. By Bishop-Gromov, $\operatorname{vol}(B(x_k, Q_k^{-\frac{1}{2}})) > \epsilon Q_k^{-\frac{n}{2}}$ at time zero, so we can parabolically rescale by Q_k and take a convergent subsequence. The limit $(M_\infty, g_\infty(\cdot))$ will be a nonflat complete ancient solution with nonnegative curvature operator, bounded curvature and $\mathcal{V}(0) > 0$. By Proposition 41.13, it cannot be κ-noncollapsed for any κ. Thus for each $\kappa > 0$, there are a point $(x_\kappa, t_\kappa) \in M_\infty \times (-\infty, 0]$ and a radius r_κ so that $|\operatorname{Rm}(x_\kappa, t_\kappa)| \le r_\kappa^{-2}$ on the time-t_κ ball $B(x_\kappa, r_\kappa)$, but $\operatorname{vol}(B(x_\kappa, r_\kappa)) < \kappa\, r_\kappa^n$. From the Bishop-Gromov inequality, $\mathcal{V}(t_\kappa) < \kappa$ for the limit solution.

We claim that $\mathcal{V}(t)$ is nonincreasing in t. To see this, we have $\frac{\operatorname{dvol}(U)}{dt} = \int_U R\, dV \ge 0$ for any domain $U \subset M_\infty$. Also, as $R \le 2$ on $M_\infty \times (-\infty, 0]$, Corollary 27.16 gives that distances on M_∞ decrease at most linearly in t, which implies the claim.

Thus $\mathcal{V}(0) = 0$ for the limit solution, which is a contradiction. $\square$

45. I.11.6. Curvature bounds for Ricci flow solutions with nonnegative curvature operator, assuming a lower volume bound

In this section we show that for a Ricci flow solution with nonnegative curvature operator, a lower bound on the volume of a ball implies an earlier upper curvature bound on a slightly smaller ball. This will be used in Section 54.

Corollary 45.1. *(cf. Corollary I.11.6) For every $w > 0$, there are $B = B(w) < \infty$, $C = C(w) < \infty$ and $\tau_0 = \tau_0(w) > 0$ with the following properties.*
(a) Take $t_0 \in [-r_0^2, 0)$. Suppose that we have a (not necessarily complete) Ricci flow solution $(M, g(\cdot))$, defined for $t \in [t_0, 0]$, so that at time zero the metric ball $B(x_0, r_0)$ has compact closure. Suppose that for each $t \in [t_0, 0]$, $g(t)$ has nonnegative curvature operator and $\mathrm{vol}(B_t(x_0, r_0)) \geq w r_0^n$. Then

$$R(x,t) \leq C r_0^{-2} + B(t - t_0)^{-1} \tag{45.2}$$

whenever $\mathrm{dist}_t(x, x_0) \leq \frac{1}{4} r_0$.
(b) Suppose that we have a (not necessarily complete) Ricci flow solution $(M, g(\cdot))$, defined for $t \in [-\tau_0 r_0^2, 0]$, so that at time zero the metric ball $B(x_0, r_0)$ has compact closure. Suppose that for each $t \in [-\tau_0 r_0^2, 0]$, $g(t)$ has nonnegative curvature operator. If we assume a time-zero volume bound $\mathrm{vol}(B_0(x_0, r_0)) \geq w r_0^n$ then

$$R(x,t) \leq C r_0^{-2} + B(t + \tau_0 r_0^2)^{-1} \tag{45.3}$$

whenever $t \in [-\tau_0 r_0^2, 0]$ and $\mathrm{dist}_t(x, x_0) \leq \frac{1}{4} r_0$.

Remark 45.4. The statement in [51, Corollary 11.6(a)] does not have any constraint on t_0. In our proof we seem to need that $-t_0 \leq c r_0^2$ for some arbitrary but fixed constant $c < \infty$. (The statement $R(\overline{x}, \overline{t}) > C + B(\overline{t} - t_0)^{-1}$ in [51, Proof of Corollary 11.6(a)] is the issue.) For simplicity we take $-t_0 \leq r_0^2$. This point does not affect the proof of Corollary 45.1(b), which is what ends up getting used.

Proof. For part (a), we can assume that $r_0 = 1$. Given $B, C > 0$, suppose that $g(\cdot)$ is a Ricci flow solution for $t \in [t_0, 0]$ that satisfies the hypotheses of the corollary, with $R(x,t) > C + B(t - t_0)^{-1}$ for some (x,t) satisfying $\mathrm{dist}_t(x, x_0) \leq \frac{1}{4}$. Following the notation of the proof of Theorem 30.1, except changing the A of Theorem 30.1 to $\widehat{A}$, put $\widehat{A} = \lambda C^{\frac{1}{2}}$ and $\alpha = \min(\lambda^2 C^{\frac{1}{2}}, B)$, where we will take λ to be a sufficiently small number that only depends on n. Put

$$M_\alpha = \{(x', t') : R(x', t') \geq \alpha (t' - t_0)^{-1}\}. \tag{45.5}$$

Clearly $(x,t) \in M_\alpha$.

We first go through the analog of the proof of Lemma 31.1. We claim that there is some $(\overline{x}, \overline{t}) \in M_\alpha$, with $\overline{t} \in (t_0, 0]$ and $\mathrm{dist}_{\overline{t}}(\overline{x}, x_0) \leq \frac{1}{3}$, such that $R(x', t') \leq 2Q = 2R(\overline{x}, \overline{t})$ whenever $(x', t') \in M_\alpha$, $t' \in (t_0, \overline{t}]$ and $\mathrm{dist}_{t'}(x', x_0) \leq \mathrm{dist}_{\overline{t}}(\overline{x}, x_0) + \widehat{A} Q^{-\frac{1}{2}}$. Put $(x_1, t_1) = (x, t)$. Inductively, if we cannot take (x_k, t_k) for $(\overline{x}, \overline{t})$ then there is some $(x_{k+1}, t_{k+1}) \in M_\alpha$ with $t_{k+1} \in (t_0, t_k]$, $R(x_{k+1}, t_{k+1}) > 2R(x_k, t_k)$ and $\mathrm{dist}_{t_{k+1}}(x_{k+1}, x_0) \leq \mathrm{dist}_{t_k}(x_k, x_0) +$

$\widehat{A}R(x_k,t_k)^{-\frac{1}{2}}$. As the process must terminate, we end up with $(\overline{x},\overline{t})$ satisfying

$$\mathrm{dist}_{\overline{t}}(\overline{x},x_0) \;\le\; \frac{1}{4} \;+\; \frac{1}{1-\sqrt{1/2}}\,\widehat{A}\,R(x,t)^{-\frac{1}{2}} \;\le\; \frac{1}{3} \tag{45.6}$$

if λ is sufficiently small.

Next, we go through the analog of the proof of Lemma 32.1. As in the proof of Lemma 32.1, $R(x',t') \le 2R(\overline{x},\overline{t})$ whenever $\overline{t}-\frac{1}{2}\alpha Q^{-1} \le t' \le \overline{t}$ and $\mathrm{dist}_{t'}(x',x_0) \le \mathrm{dist}_{\overline{t}}(\overline{x},x_0)+\widehat{A}Q^{-\frac{1}{2}}$. We claim that the time-$\overline{t}$ ball $B(x_0,\mathrm{dist}_{\overline{t}}(\overline{x},x_0)+\frac{1}{10}\widehat{A}Q^{-1/2})$ is contained in the time-t' ball $B(x_0,\mathrm{dist}_{\overline{t}}(\overline{x},x_0)+\widehat{A}Q^{-\frac{1}{2}})$. To see this, we apply Lemma 27.8 with $r_0 = \frac{1}{2}Q^{-1/2}$ to give

$$\mathrm{dist}_t(x_0,\overline{x}) \;-\; \mathrm{dist}_{\overline{t}}(x_0,\overline{x}) \;\le\; \mathrm{const.}(n)\,\alpha Q^{-1/2} \;\le\; \lambda\,\mathrm{const.}(n)\widehat{A}Q^{-1/2}. \tag{45.7}$$

If λ is sufficiently small then the claim follows. The argument also shows that it is consistent to use the curvature bound when applying Lemma 27.8.

Hence $R(x',t') \le 2R(\overline{x},\overline{t})$ whenever $\overline{t}-\frac{1}{2}\alpha Q^{-1} \le t' \le \overline{t}$ and $\mathrm{dist}_{\overline{t}}(x',x_0) \le \mathrm{dist}_{\overline{t}}(\overline{x},x_0)+\frac{1}{10}\widehat{A}Q^{-1/2}$. It follows that $R(x',t') \le 2R(\overline{x},\overline{t})$ whenever $\overline{t}-\frac{1}{2}\alpha Q^{-1} \le t' \le \overline{t}$ and $\mathrm{dist}_{\overline{t}}(x',\overline{x}) \le \frac{1}{10}\widehat{A}Q^{-1/2}$. This shows that there is an $A' = A'(B,C)$, which goes to infinity as $B,C \to \infty$, so that $R(x',t') \le 2R(\overline{x},\overline{t})$ whenever $\overline{t} - A'Q^{-1} \le t' \le \overline{t}$ and $\mathrm{dist}_{\overline{t}}(x',\overline{x}) \le A'Q^{-1/2}$.

Now suppose that Corollary 45.1(a) is not true. Fixing $w>0$, for any sequences $\{B_k\}_{k=1}^\infty$ and $\{C_k\}_{k=1}^\infty$ going to infinity and for each k, there is a Ricci flow solution $g_k(\cdot)$ which satisfies the hypotheses of the corollary but for which $R(x_k,t_k) \ge C_k + B_k(t_k-t_{0,k})^{-1}$ for some point (x_k,t_k) satisfying $\mathrm{dist}_{t_k}(x_k,x_{0,k}) \le \frac{1}{4}$. We can assume that $\lambda^2 C_k^{\frac{1}{2}} \ge B_k$. From the preceding discussion, there is a sequence $A'_k \to \infty$ and points $(\overline{x}_k,\overline{t}_k)$ with $\mathrm{dist}_{\overline{t}_k}(\overline{x}_k,x_{0,k}) \le \frac{1}{3}$ so that $R(x'_k,t'_k) \le 2R(\overline{x}_k,\overline{t}_k)$ whenever $\overline{t}_k - A'_kQ_k^{-1} \le t'_k \le \overline{t}_k$ and $\mathrm{dist}_{\overline{t}_k}(x'_k,\overline{x}_k) \le A'_kQ_k^{-1/2}$, where

$$Q_k \;=\; R(\overline{x}_k,\overline{t}_k) \;\ge\; B_k\,(\overline{t}_k-t_{0,k})^{-1} \;\ge\; B_k. \tag{45.8}$$

By Corollary 44.1, for any $\epsilon>0$ there is some $A = A(\epsilon) < \infty$ so that for large k,

$$\mathrm{vol}(B(\overline{x}_k,A/\sqrt{Q_k})) \;\le\; \epsilon(A/\sqrt{Q_k})^n \tag{45.9}$$

at time zero. By the Bishop-Gromov inequality, $\mathrm{vol}(B(\overline{x}_k,1)) \le \epsilon$ for large k, since $Q_k\to\infty$. If we took ϵ sufficiently small from the beginning then we would get a contradiction to the fact that

$$\mathrm{vol}(B(\overline{x}_k,1)) \;\ge\; \mathrm{vol}\left(B\left(x_{0,k},\frac{2}{3}\right)\right) \;\ge\; \left(\frac{2}{3}\right)^n \mathrm{vol}(B(x_{0,k},1)) \;\ge\; \left(\frac{2}{3}\right)^n w. \tag{45.10}$$

For part (b), the idea is to choose the parameter τ_0 sufficiently small so that we will still have the estimate $\mathrm{vol}(B(x_0,r_0)) \ge 5^{-n}\,w\,r_0^n$ for the time-t ball $B(x_0,r_0)$ when $t\in[-\tau_0r_0^2,0]$, and so we can apply part (a) with w replaced by $\frac{w}{5}$. The value of τ_0 will emerge from the proof. More precisely, putting $r_0 = 1$ and with a given τ_0, let τ be the largest number in $[0,\tau_0]$ so that the time-t ball $B(x_0,1)$ satisfies $\mathrm{vol}(B(x_0,1)) \ge 5^{-n}\,w$ whenever $t\in[-\tau,0]$. If $\tau < \tau_0$ then at time $-\tau$, we have $\mathrm{vol}(B(x_0,1)) = 5^{-n}\,w$. The conclusion of part (a) holds in the sense that

$$R(x,t) \;\le\; C(5^{-n}w) \;+\; B(5^{-n}w)(t+\tau)^{-1} \tag{45.11}$$

whenever $t \in [-\tau, 0]$ and $\mathrm{dist}_t(x, x_0) \le \frac{1}{4}$. Lemma 27.8, along with (45.11), implies that the time-$(-\tau)$ ball $B(x_0, \frac{1}{4})$ contains the time-0 ball $B(x_0, \frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}))$. From the nonnegative curvature, the time-$(-\tau)$ volume of the first ball is at least as large as the time-0 volume of the second ball. Then

$$\begin{aligned} 5^{-n}\, w &= \mathrm{vol}(B(x_0, 1)) \ge \mathrm{vol}(B(x_0, \frac{1}{4})) \\ &\ge \mathrm{vol}(B(x_0, \frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}))) \\ &\ge (\frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}))^n \, \mathrm{vol}(B(x_0, 1)) \\ &\ge (\frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}))^n \, w, \end{aligned} \tag{45.12}$$

where the balls on the top line of (45.12) are at time-$(-\tau)$, and the other balls are at time-0. Thus $\frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}) \le \frac{1}{5}$. This contradicts our assumption that $\tau < \tau_0$ provided that $\frac{1}{4} - 10(n-1)(\tau_0\sqrt{C} + 2\sqrt{B\tau_0}) = \frac{1}{5}$. □

Finally, we give a version of Corollary 45.1(b) where instead of assuming a nonnegative curvature operator, we assume that the curvature operator in the time-dependent ball of radius r_0 around x_0 is bounded below by $-r_0^{-2}$.

Corollary 45.13. *(cf. end of Section I.11.6) For every $w > 0$, there are $B = B(w) < \infty$, $C = C(w) < \infty$ and $\tau_0 = \tau_0(w) > 0$ with the following property. Suppose that we have a (not necessarily complete) Ricci flow solution $(M, g(\cdot))$, defined for $t \in [-\tau_0 r_0^2, 0]$, so that at time zero the metric ball $B(x_0, r_0)$ has compact closure. Suppose that for each $t \in [-\tau_0 r_0^2, 0]$, the curvature operator in the time-t ball $B(x_0, r_0)$ is bounded below by $-r_0^{-2}$. If we assume a time-zero volume bound* $\mathrm{vol}(B_0(x_0, r_0)) \ge w r_0^n$ *then*

$$R(x,t) \le C r_0^{-2} + B(t + \tau_0 r_0^2)^{-1} \tag{45.14}$$

whenever $t \in [-\tau_0 r_0^2, 0]$ and $\mathrm{dist}_t(x, x_0) \le \frac{1}{4} r_0$.

Proof. The blowup argument goes through as before. The only real difference is that the volume of the time-$(-\tau)$ ball $B(x_0, \frac{1}{4})$ will be at least $e^{-\mathrm{const.}\, \tau r_0^{-2}}$ times the volume of the time-0 ball $B(x_0, \frac{1}{4} - 10(n-1)(\tau\sqrt{C} + 2\sqrt{B\tau}))$. □

46. I.11.7. Compactness of the space of three-dimensional κ-solutions

In this section we prove a compactness result for the space of three-dimensional κ-solutions. The three-dimensionality assumption is used to show that the limit solution has bounded curvature.

If a three-dimensional κ-solution M is compact then it is diffeomorphic to a quotient of S^3 or $\mathbb{R} \times S^2$, as it has nonnegative curvature and is nonflat. If its asymptotic soliton (see Section 39) is also closed then M is a quotient of the round S^3 or $\mathbb{R} \times S^2$. There are κ-solutions on S^3 and $\mathbb{R}P^3$ with noncompact asymptotic soliton; see [52, Section 1.4]. They are not isometric to the round metric; this corrects the statement in the first paragraph of [51, Section 11.7].

Theorem 46.1. *(cf. Theorem I.11.7) Given $\kappa > 0$, the set of oriented three-dimensional κ-solutions is compact modulo scaling. That is, from any sequence of such solutions and points $(x_k, 0)$, after appropriate dilations we can extract a smoothly converging subsequence that satisfies the same conditions.*

Proof. If $(M_k, (x_k, 0), g_k(\cdot))$ is a sequence of such κ-solutions with $R(x_k, 0) = 1$ then parts 1 and 3 of Corollary 42.1 imply there is a subsequence that converges to an ancient solution $(M_\infty, (x_\infty, 0), g_\infty(\cdot))$ which has nonnegative curvature operator and is κ-noncollapsed. The remaining issue is to show that it has bounded curvature. Note that $R_t \geq 0$ since $g_\infty(\cdot)$ is a limit of a sequence of Ricci flows satisfying $R_t \geq 0$. Hence it is enough to show that $(M_\infty, g(0))$ has bounded scalar curvature.

If not, there is a sequence of points y_i going to infinity in M_∞ such that $R(y_i, 0) \to \infty$ and $R(y, 0) \leq 2R(y_i, 0)$ for $y \in B(y_i, A_i R(y_i, 0)^{-\frac{1}{2}})$, where $A_i \to \infty$; compare [33, Lemma 22.2]. Using the κ-noncollapsing, a subsequence of the rescalings $(M_\infty, y_i, R(y_i, 0) g_\infty)$ will converge to a limit manifold N_∞. As in the proof of Proposition 41.13 from Appendix G, N_∞ will split off a line. By Corollary 40.1 or Section 43, N_∞ must be the standard solution on $\mathbb{R} \times S^2$. Thus $(M, g(0))$ contains a sequence D_i of neck regions, with their cross-sectional radii tending to zero as $i \to \infty$.

Note that M has to be 1-ended. Otherwise, it would contain a line, and would therefore have to split off a line isometrically [18, Theorem 8.17]. But then M, the product of a line and a surface, could not have neck regions with cross-sections tending to zero.

From the theory of nonnegatively curved manifolds [18, Chapter 8.5], there is an exhaustion $M = \bigcup_{t \geq 0} C_t$ by nonempty totally convex compact sets C_t so that $(t_1 \leq t_2) \Rightarrow (C_{t_1} \subset C_{t_2})$, and

$$C_{t_1} = \{q \in C_{t_2} : \operatorname{dist}(q, \partial C_{t_2}) \geq t_2 - t_1\}. \tag{46.2}$$

Now consider a neck region D which is close to a cylinder. Note by triangle comparison – or simply because the distance function in D is close to that of a product metric – any minimizing geodesic segment $\gamma \subset D$ of length large compared to cross-sectional radius of D must be nearly orthogonal to the cross-section. It follows from this and (46.2) that if $t > 0$ and ∂C_t contains a point $p \in D$ such that $d(p, \partial D)$ is large compared to the cross-section of D, then $\partial C_t \cap D$ is an approximate 2-sphere cross-section of D. Fix such a neck region D_0 and let C_{t_0} be the corresponding convex set. As M has one end, ∂C_{t_0} has only one connected component, namely the approximate 2-sphere cross-section.

For all $t > t_0$, there is a distance-nonincreasing retraction $r : C_t \to C_{t_0}$ which maps $C_t - C_{t_0}$ onto ∂C_{t_0} [60]. Let D be a neck region with a very small cross-section and let C_t be a convex set so that ∂C_t intersects D in an approximate 2-sphere cross-section. Then ∂C_t consists entirely of this approximate cross-section. The restriction of r to ∂C_t is distance-nonincreasing, but will map the 2-sphere ∂C_t onto the 2-sphere ∂C_{t_0}. This is a contradiction. $\square$

Remark 46.3. The statement of [51, Theorem 11.7] is about noncompact κ-solutions but the proof works whether the solutions are compact or noncompact.

Remark 46.4. One may wonder where we have used the fact that we have a Ricci flow solution, i.e. whether the curvature is bounded for any κ-noncollapsed Riemannian 3-manifold with nonnegative sectional curvature. Following the above argument, we could again split off a line in a rescaling around high-curvature points. However, we would not necessarily know that the ensuing nonnegatively-curved surface is compact. (*A priori*, it could be a smoothed-out cone, for example.) In the case of a Ricci flow, the compactness comes from Corollary 40.1 or Section 43.

Corollary 46.5. *Let $(M, g(\cdot))$ be a 3-dimensional κ-solution. Then any asymptotic soliton constructed as in Section 39 is also a κ-solution.*

47. I.11.8. Necklike behavior at infinity of a three-dimensional κ-solution - weak version

The next corollary says that outside of a compact region, any oriented noncompact three-dimensional κ-solution looks necklike (after rescaling). In this section we give a simple argument to prove the corollary, except for a diameter bound on the compact region. In the next section we give an argument that also proves the diameter bound.

More information on three-dimensional κ-solutions is in Section 59.

Definition 47.1. Fix $\epsilon > 0$. Let $(M, g(\cdot))$ be an oriented three-dimensional κ-solution. We say that a point $x_0 \in M$ is the *center of an ϵ-neck* if the solution $g(\cdot)$ in the set $\{(x,t) \, : \, -(\epsilon Q)^{-1} < t \le 0, \mathrm{dist}_0(x, x_0)^2 < (\epsilon Q)^{-1}\}$, where $Q = R(x_0, 0)$, is, after scaling with the factor Q, ϵ-close in some fixed smooth topology to the corresponding subset of the evolving round cylinder (having scalar curvature one at time zero). (See Definition 58.1 below for a more precise statement.)

We let M_ϵ denote the points in M that are not centers of ϵ-necks.

Corollary 47.2. *(cf. Corollary I.11.8) For any $\epsilon > 0$, there exists $C = C(\epsilon, \kappa) > 0$ such that if $(M, g(\cdot))$ is an oriented noncompact three-dimensional κ-solution then*
1. M_ϵ is compact with $\mathrm{diam}(M_\epsilon) \le CQ^{-\frac{1}{2}}$ and
2. $C^{-1}Q \le R(x, 0) \le CQ$ whenever $x \in M_\epsilon$,
where $Q = R(x_0, 0)$ for some $x_0 \in \partial M_\epsilon$.

Proof. We prove here the claims of Corollary 47.2, except for the diameter bound. In the next section we give another argument which also proves the diameter bound.

We claim first that M_ϵ is compact. Suppose not. Then there is a sequence of points $x_k \in M_\epsilon$ going to infinity. Fix a basepoint $x_0 \in M$. Then $R(x_0)\ \mathrm{dist}_0^2(x_0, x_k) \to \infty$. By part 5 of Corollary 42.1, $R(x_k)\ \mathrm{dist}_0^2(x_0, x_k) \to \infty$. Rescaling around $(x_k, 0)$ to make its scalar curvature one, we can use Theorem 46.1 to extract a convergent subsequence (M_∞, x_∞). As in the proof of Proposition 41.13, we can say that (M_∞, x_∞) splits off a line. Hence for large k, x_k is the center of an ϵ-neck, which is a contradiction.

Next we claim that for any ϵ, there exists $C = C(\epsilon, \kappa) > 0$ such that if $g_{ij}(t)$ is a κ-solution then for any point $x \in M_\epsilon$, there is a point $x_0 \in \partial M_\epsilon$ such that $\mathrm{dist}_0(x, x_0) \le CQ^{-1/2}$ and $C^{-1}Q \le R(x, 0) \le CQ$, where $Q = R(x_0, 0)$.

If not then there is a sequence $\{M_i\}_{i=1}^{\infty}$ of κ-solutions along with points $x_i \in M_{i,\epsilon}$ such that for each $y_i \in \partial M_{i,\epsilon}$, we have
1. $\mathrm{dist}_0^2(x_i, y_i)\, R(y_i, 0) \geq i$ or
2. $R(y_i, 0) \geq i\, R(x_i, 0)$ or
3. $R(x_i, 0) \geq i\, R(y_i, 0)$.

Rescale the metric on M_i so that $R(x_i, 0) = 1$. From Theorem 46.1, a subsequence of the pointed spaces (M_i, x_i) will converge smoothly to a κ-solution (M_∞, x_∞). Also, $x_\infty \in M_{\infty,\epsilon}$.

Taking a subsequence, we can assume that 1. occurs for each i, or 2. occurs for each i, or 3. occurs for each i. If $M_\infty \neq M_{\infty,\epsilon}$, choose $y_\infty \in \partial M_{\infty,\epsilon}$. Then y_∞ is the limit of a subsequence of points $y_i \in \partial M_{i,\epsilon}$.

If 1. occurs for each i then $\mathrm{dist}_0^2(x_\infty, y_\infty)\, R(y_\infty, 0) = \infty$, which is impossible. If 2. occurs for each i then $R(y_\infty, 0) = \infty$, which is impossible. If 3. occurs for each i then $R(y_\infty, 0) = 0$. It follows from (F.14) that M_∞ is flat, which is impossible, as $R(x_\infty, 0) = 1$.

Hence $M_\infty = M_{\infty,\epsilon}$, i.e. no point in the noncompact ancient solution M_∞ is the center of an ϵ-neck. This contradicts the previous conclusion that $M_{\infty,\epsilon}$ is compact. □

48. I.11.8. Necklike behavior at infinity of a three-dimensional κ-solution - strong version

The following corollary is an application of the compactness result Theorem 46.1. it is a refinement of [51, Cor. I.11.8].

Corollary 48.1. *For all $\kappa > 0$, there exists an $\epsilon_0 > 0$ such that for all $0 < \epsilon < \epsilon_0$ there exists an $\alpha = \alpha(\epsilon, \kappa)$ with the property that for any κ-solution $(M, g(\cdot))$, and at any time t, precisely one of the following holds (M_ϵ denotes the set of points which are not centers of ϵ-necks at time t):*

A. $(M, g(\cdot))$ is round cylindrical flow, and so every point at every time is the center of an ϵ-neck for all $\epsilon > 0$.

B. M is noncompact, $M_\epsilon \neq \emptyset$, and for all $x, y \in M_\epsilon$, we have $R(x)d^2(x, y) < \alpha$.

C. M is compact, and there is a pair of points $x, y \in M_\epsilon$ such that $R(x)d^2(x, y) > \alpha$,

$$M_\epsilon \subset B(x, \alpha R(x)^{-\frac{1}{2}}) \cup B(y, \alpha R(y)^{-\frac{1}{2}}), \tag{48.2}$$

and there is a minimizing geodesic $\overline{xy}$ such that every $z \in M - M_\epsilon$ satisfies $R(z)d^2(z, \overline{xy}) < \alpha$.

D. M is compact and there exists a point $x \in M_\epsilon$ such that $R(x)d^2(x, z) < \alpha$ for all $z \in M$.

Lemma 48.3. *For all $\epsilon > 0$, $\kappa > 0$, there exists $\alpha = \alpha(\epsilon, \kappa)$ with the following property. Suppose $(M, g(\cdot))$ is any κ-solution, $x, y, z \in M$, and at time t we have $x, y \in M_\epsilon$ and $R(x)d^2(x, y) > \alpha$. Then at time t either $R(x)d^2(z, x) < \alpha$ or $R(y)d^2(z, y) < \alpha$ or $(R(z)d^2(z, \overline{xy}) < \alpha$ and $z \notin M_\epsilon)$.*

Proof. Pick $\epsilon > 0$, $\kappa > 0$, and suppose no such α exists. Then there is a sequence $\alpha_k \to \infty$, a sequence of κ-solutions $(M_k, g_k(\cdot))$, and sequences $x_k, y_k, z_k \in M_k$, $t_k \in \mathbb{R}$ violating the

α_k-version of the statement for all k. In particular, $x_k, y_k \in (M_k)_\epsilon$ and

(48.4) $R(x_k,t_k)d^2_{t_k}(x_k,y_k) \to \infty,\ R(x_k,t_k)d^2_{t_k}(z_k,x_k) \to \infty,$ and $R(y_k,t_k)d^2_{t_k}(z_k,y_k) \to \infty.$

Let $z'_k \in \overline{x_k y_k}$ be a point in $\overline{x_k y_k}$ nearest z_k in $(M_k, g_k(t_k))$.

We first show that $R(x_k,t_k)d^2_{t_k}(z'_k,x_k) \to \infty$. If not, we may pass to a subsequence on which $R(x_k,t_k)d^2_{t_k}(z'_k,x_k)$ remains bounded. Applying Theorem 46.1, we may pass to a subsequence and rescale by $R(x_k,t_k)$, to make the sequence $(M_k,(x_k,t_k),g_k(\cdot))$ converge to a κ-solution $(M_\infty,(x_\infty,0),g_\infty(\cdot))$, the segments $\overline{x_k y_k} \subset (M_k, g_k(t_k))$ converge to a ray $\overline{x_\infty \xi} \subset (M_\infty, g_\infty(0))$, and the segments $\overline{z'_k z_k}$ converge to a ray $\overline{z'_\infty \eta}$. Recall that the comparison angle $\widetilde{\angle}_{z'_\infty}(u,v)$ tends to the Tits angle $\partial_T(\xi,\eta)$ as $u \in \overline{z'_\infty \xi}$, $v \in \overline{z'_\infty \eta}$ tend to infinity. Since $d(z_k,z'_k) = d(z_k, \overline{x_k y_k})$ we must have $\partial_T(\xi,\eta) \geq \frac{\pi}{2}$. Now consider a sequence $u_k \in \overline{z'_\infty \xi}$ tending to infinity. By Theorem 46.1, part 5 of Corollary 42.1, and the remarks about Alexandrov spaces in Appendix G, if we rescale $(M_\infty,(u_k,0),g_\infty(\cdot))$ by $R(u_k,0)$, we get round cylindrical flow as a limit. When k is sufficiently large, we may find an almost product region $D \subset (M_\infty, g_\infty(\cdot))$ containing u_k which is disjoint from $\overline{z'_\infty \eta}$, and whose cross-section $\Sigma \times \{0\} \subset \Sigma \times (-1,1) \simeq D$ intersects the ray $\overline{z'_\infty \xi}$ transversely at a single point. This implies that $\Sigma \times \{0\}$ separates the two ends of $\overline{z'_\infty \xi} \cup \overline{z'_\infty \eta}$ from each other; hence M_∞ is two-ended, and $(M_\infty, g_\infty(\cdot))$ is round cylindrical flow. This contradicts the assumption that x_k is not the center of an ϵ-neck. Hence $R(x_k,t_k)d^2_{t_k}(z'_k,x_k) \to \infty$, and similar reasoning shows that $R(y_k,t_k)d^2_{t_k}(z'_k,y_k) \to \infty$.

By part 5 of Corollary 42.1, we therefore have $R(z'_k,t_k)d^2_{t_k}(z'_k,x_k) \to \infty$ and $R(z'_k,t_k)d^2_{t_k}(z'_k,y_k) \to \infty$. Rescaling the sequence $(M_k,(z'_k,t_k),g_k(\cdot))$ by $R(z'_k,t_k)$, we get convergence to round cylindrical flow (since any limit flow contains a line), and $\overline{z'_k z_k}$ subconverges to a segment orthogonal to the $\mathbb{R}$-factor, which implies that $R(z'_k,t_k)d^2_{t_k}(z_k,z'_k)$ is bounded and z_k is the center of an ϵ-neck for large k. This contradicts our assumption that the α_k-version of the lemma is violated for each k. □

Proof of Corollary 48.1. Let $(M, g(\cdot))$ be a κ-solution, and $\epsilon > 0$.

Case 1: Every $x \in (M, g(t))$ is the center of an ϵ-neck. In this case, if $\epsilon > 0$ is sufficiently small, M fibers over a 1-manifold with fiber S^2. If the 1-manifold is homeomorphic to $\mathbb{R}$, then M has two ends, which implies that the flow $(M, g(\cdot))$ is an evolving round cylinder. If the base of the fibration were a circle, then the universal cover $(\tilde{M}, \tilde{g}(t))$ would split off a line, which would imply that the universal covering flow would be a round cylindrical flow; but this would violate the κ-noncollapsed assumption at very negative times. Thus A holds in this case.

Case 2: There exist $x, y \in M_\epsilon$ such that $R(x)d^2(x,y) > \alpha$. By Lemma 48.3 and Corollary 42.1 part 5, for all $z \in M - \left(B(x, \alpha R(x)^{-\frac{1}{2}}) \cup B(y, \alpha R(y)^{-\frac{1}{2}})\right)$, we have $R(z)d^2(z, \overline{xy}) < \alpha$ and $z \notin M_\epsilon$. This implies (again by Corollary 42.1 part 5) that there exists a $\gamma = \gamma(\epsilon, \kappa)$ such that for every $z \in M$ there is a $z' \in \overline{xy}$ for which $R(z')d^2(z',z) < \gamma$, which means that M must be compact, and C holds.

Case 3: $M_\epsilon \neq \emptyset$, and for all $x, y \in M_\epsilon$, we have $R(x)d^2(x,y) < \alpha$. If M is noncompact then we are in case B and are done, so assume that M is compact. Pick $x \in M_\epsilon$, and suppose $z \in M$ maximizes $R(x)d^2(\cdot, x)$. If $R(x)d^2(z,x) \geq \alpha$, then z is the center of an ϵ-neck, and we may look at the cross-section Σ of the neck region. If Σ separates M, then when $\epsilon > 0$ is sufficiently small, we get a contradiction to the assumption that z maximizes $R(x)d^2(x, \cdot)$. Hence Σ cannot separate M, and there is a loop passing through x which intersects Σ transversely at one point. It follows that the universal covering flow $(\tilde{M}, \tilde{g}(\cdot))$ is cylindrical flow, a contradiction. Hence $R(x)d^2(x,z) < \alpha$ for all $z \in M$, so D holds. □

49. More properties of κ-solutions

In this section we prove some additional properties of κ-solutions. In particular, Corollary 49.2 implies that if z lies in a geodesic segment η in a κ-solution M and if the endpoints of η are sufficiently far from z (relative to $R(z)^{-\frac{1}{2}}$) then $z \notin M_\epsilon$. The results of this section will be used in the proof of Theorem 52.7.

Proposition 49.1. *For all $\kappa > 0$, $\alpha > 0$, $\theta > 0$, there exists a $\beta(\kappa, \alpha, \theta) < \infty$ such that if $(M, g(t))$ is a time slice of a κ-solution, $x, y_1, y_2 \in M$, $R(x)d^2(x, y_i) > \beta$ for $i = 1, 2$, and $\widetilde{\angle}_x(y_1, y_2) \geq \theta$, then (a) x is the center of an α-neck, and (b) $\widetilde{\angle}_x(y_1, y_2) \geq \pi - \alpha$.*

Proof. The proof of this is similar to the first part of the proof of Lemma 48.3. Note that when α is small, then after enlarging β if necessary, the neck region around x will separate y_1 from y_2; this implies (b). □

Corollary 49.2. *For all $\kappa > 0$, $\epsilon > 0$, there exists a $\rho = \rho(\kappa, \epsilon)$ such that if $(M, g(t))$ is a time slice of a κ-solution, $\eta \subset (M, g(t))$ is a minimizing geodesic segment with endpoints y_1, y_2, $z \in M$, $z' \in \eta$ is a point in η nearest z, and $R(z')d^2(z', y_i) > \rho$ for $i = 1, 2$, then z, z' are centers of ϵ-necks, and* $\max(R(z)d^2(z, z'), R(z')d^2(z, z')) < 4\pi^2$.

Proof. Pick $\epsilon' > 0$. Under the assumptions, if

$$\min(R(z')d^2(z', y_1), R(z')d^2(z', y_2)) \tag{49.3}$$

is sufficiently large, we can apply the preceding proposition to the triple z', y_1, y_2, to conclude that z' is the center of an ϵ'-neck. Since the shortest segment from z to z' is orthogonal to η, when ϵ' is small enough the segment $\overline{zz'}$ will lie close to an S^2 cross-section in the approximating round cylinder, which gives $R(z')d^2(z, z') \lesssim 2\pi^2$. □

50. I.11.9. Getting a uniform value of κ

Proposition 50.1. *There is a $\kappa_0 > 0$ so that if $(M, g(\cdot))$ is an oriented three-dimensional κ-solution, for some $\kappa > 0$, then it is a κ_0-solution or it is a quotient of the round shrinking S^3.*

Proof. Let $(M, g(\cdot))$ be a κ-solution. Suppose that for some $\kappa' > 0$, the solution is κ'-collapsed at some scale. After rescaling, we can assume that there is a point $(x_0, 0)$ so that $|\operatorname{Rm}(x,t)| \leq 1$ for all (x,t) satisfying $\operatorname{dist}_0(x, x_0) < 1$ and $t \in [-1, 0]$, with $\operatorname{vol}(B_0(x_0, 1)) < \kappa'$. Let $\widetilde{V}(t)$ denote the reduced volume as a function of $t \in (-\infty, 0]$, as defined using curves

from $(x_0, 0)$. It is nondecreasing in t. As in the proof of Theorem 26.2, there is an estimate $\widetilde{V}(-\kappa') \leq 3(\kappa')^{3/2}$. Take a sequence of times $t_i \to -\infty$. For each t_i, choose $q_i \in M$ so that $l(q_i, t_i) \leq \frac{3}{2}$. From the proof of Proposition 39.1, for all $\epsilon > 0$ there is a $\delta > 0$ such that $l(q, t)$ does not exceed δ^{-1} whenever $t \in [t_i, t_i/2]$ and $\mathrm{dist}^2_{t_i}(q, q_i) \leq \epsilon^{-1} t_i$. Given the monotonicity of $\widetilde{V}$ and the upper bound on $l(q,t)$, we obtain an upper bound on the volume of the time-t_i ball $B(q_i, \sqrt{t_i/\epsilon})$ of the form const. $t_i^{3/2}\, e^{\delta^{-1}}\, (\kappa')^{3/2}$.

On the other hand, from Proposition 39.1, a subsequence of the rescalings of the ancient solution around (q_i, t_i) converges to a nonflat gradient shrinking soliton. If the gradient shrinking soliton is compact then it must be a quotient of the round shrinking S^3 [35]. Otherwise, Corollary 51.22 says that if the gradient shrinking soliton is noncompact then it must be an evolving cylinder or its $\mathbb{Z}_2$-quotient. Fixing ϵ, this gives a lower bound on $\mathrm{vol}(B_{t_i}(q_i, \sqrt{t_i/\epsilon}))$ in terms of the noncollapsing constants of the evolving cylinder and its $\mathbb{Z}_2$-quotient. Hence there is a universal constant κ_0 so that if $\kappa' < \kappa_0$ then we obtain a contradiction to the assumption of κ'-collapsing. □

Remark 50.2. The hypotheses of Corollary 51.22 assume a global upper bound on the sectional curvature of any time slice, which in the n-dimensional case is not *a priori* true for the asymptotic soliton of Proposition 39.1. However, in our 3-dimensional case, the argument of Theorem 46.1 shows that there is such an upper bound.

51. II.1.2. Three-dimensional noncompact κ-noncollapsed gradient shrinkers are standard

In this section we show that any complete oriented 3-dimensional noncompact κ-noncollapsed gradient shrinking soliton with bounded nonnegative curvature is either the evolving round cylinder $\mathbb{R} \times S^2$ or its $\mathbb{Z}_2$-quotient.

The basic example of a gradient shrinking soliton is the metric on $\mathbb{R} \times S^2$ which gives the 2-sphere a radius of $\sqrt{-2t}$ at time $t \in (-\infty, 0)$. With coordinates (s, θ) on $\mathbb{R} \times S^2$, the function f is given by $f(t, s, \theta) = -\frac{s^2}{4t}$.

Lemma 51.1. *(cf. Lemma of II.1.2) There is no complete oriented 3-dimensional noncompact κ-noncollapsed gradient shrinking soliton with bounded positive sectional curvature.*

Proof. The idea of the proof is to show that the soliton has the qualitative features of a shrinking cylinder, and then to get a contradiction to the assumption of positive sectional curvature.

Applying ∇_i to the gradient shrinker equation

$$\nabla_i \nabla_j f + R_{ij} + \frac{1}{2t} g_{ij} = 0 \tag{51.2}$$

gives

$$\triangle \nabla_j f + \nabla_i R_{ij} = 0. \tag{51.3}$$

As $\nabla_i R_{ij} = \frac{1}{2} \nabla_j R$ and $\triangle \nabla_j f = \nabla_j \triangle f + R_{jk} \nabla_k f = \nabla_j \left(-R - \frac{n}{2t}\right) + R_{jk} \nabla_k f$, we obtain

$$\nabla_i R = 2\, R_{ij}\, \nabla_j f. \tag{51.4}$$

Fix a basepoint $x_0 \in M$ and consider a normalized minimal geodesic $\gamma : [0, \overline{s}] \to M$ in the time -1 slice with $\gamma(0) = x_0$. Put $X(s) = \frac{d\gamma}{ds}$. As in the proof of Lemma 27.8, $\int_0^{\overline{s}} \operatorname{Ric}(X,X)\, ds \le$ const. for some constant independent of $\overline{s}$. If $\{Y_i\}_{i=1}^3$ are orthonormal parallel vector fields along γ then

$$(51.5)\qquad \left(\int_0^{\overline{s}} |\operatorname{Ric}(X,Y_1)|\, ds \right)^2 \le \overline{s} \int_0^{\overline{s}} |\operatorname{Ric}(X,Y_1)|^2\, ds \le \overline{s} \sum_{i=1}^3 \int_0^{\overline{s}} |\operatorname{Ric}(X,Y_i)|^2\, ds.$$

Thinking of Ric as a self-adjoint linear operator on TM, $\sum_{i=1}^3 |\operatorname{Ric}(X,Y_i)|^2 = \langle X, \operatorname{Ric}^2 X \rangle$. In terms of a pointwise orthonormal frame $\{e_i\}$ of eigenvectors of Ric, with eigenvalues λ_i, write $X = \sum_{i=1}^3 X_i e_i$. Then

$$(51.6)\qquad \langle X, \operatorname{Ric}^2 X\rangle = \sum_{i=1}^3 \lambda_i^2 X_i^2 \le (\sum_{i=1}^3 \lambda_i)(\sum_{i=1}^3 \lambda_i X_i^2) = R \cdot \operatorname{Ric}(X,X).$$

Hence

$$(51.7)\qquad \left(\int_0^{\overline{s}} |\operatorname{Ric}(X,Y_1)|\, ds \right)^2 \le (\sup_M R)\, \overline{s} \int_0^{\overline{s}} \operatorname{Ric}(X,X)\, ds \le \text{const. } \overline{s}.$$

Multiplying (51.2) by $X^i X^j$ and summing gives $\frac{d^2 f(\gamma(s))}{ds^2} + \operatorname{Ric}(X,X) - \frac{1}{2} = 0$. Then

$$(51.8)\qquad \frac{df(\gamma(s))}{ds}\Big|_{s=\overline{s}} = \frac{df(\gamma(s))}{ds}\Big|_{s=0} + \frac{1}{2}\overline{s} - \int_0^{\overline{s}} \operatorname{Ric}(X,X)\, ds \ge \frac{1}{2}\overline{s} - \text{const.}$$

This implies that there is a compact subset of M outside of which f has no critical points.

If Y is a unit vector field perpendicular to X then multiplying (51.2) by $X^i Y^j$ and summing gives $\frac{d}{ds}(Y \cdot f)(\gamma(s)) + \operatorname{Ric}(X,Y) = 0$. Then

$$(51.9)\qquad (Y \cdot f)(\gamma(\overline{s})) = (Y \cdot f)(\gamma(0)) - \int_0^{\overline{s}} \operatorname{Ric}(X,Y)\, ds$$

and

$$(51.10)\qquad |(Y \cdot f)(\gamma(\overline{s}))| \le \text{const.}(\sqrt{\overline{s}} + 1).$$

For large $\overline{s}$, $|(Y \cdot f)(\gamma(\overline{s}))|$ is small compared to $(X \cdot f)(\gamma(\overline{s}))$. This means that as one approaches infinity, the gradient of f becomes more and more parallel to the gradient of the distance function from x_0, where by the latter we mean the vectors X that are tangent to minimal geodesics.

The gradient flow of f is given by the equation

$$(51.11)\qquad \frac{dx}{du} = (\nabla f)(x).$$

Then along a flowline, equation (51.4) implies that

$$(51.12)\qquad \frac{dR(x)}{du} = \left\langle \nabla R, \frac{dx}{du} \right\rangle = 2 \operatorname{Ric}(\nabla f, \nabla f).$$

In particular, outside of a compact set, R is strictly increasing along the flowlines. Put $\overline{R} = \limsup_{x\to\infty} R$. Take points x_α tending toward infinity, with $R(x_\alpha) \to \overline{R}$. Putting

$r_\alpha = \sqrt{\mathrm{dist}_{-1}(x_0, x_\alpha)}$, we have $\frac{r_\alpha}{\mathrm{dist}_{-1}(x_0,x_\alpha)} \to 0$ and $R(x_\alpha)\, r_\alpha^2 \to \infty$. Then the argument of the proof of Proposition 41.13 shows that any convergent subsequence of the rescalings around $(x_\alpha, -1)$ splits off a line. Hence the limit is a shrinking round cylinder with scalar curvature $\overline{R}$ at time -1. Because our original solution exists up to time zero, we must have $\overline{R} \le 1$. Now equation (C.13) says that the Ricci flow is given by $g(-t) = -t\,\phi_t^* g(-1)$, where ϕ_t is the flow generated by ∇f. It follows that $\inf_{x\in M} R(x,t) = Ct^{-1}$ for some $C > 0$. That is, the curvature blows up uniformly as $t \to 0$. Comparing this with the singularity time of the shrinking round cylinder implies that $\overline{R} = 1$. Performing a similar argument with any sequence of x_α's tending toward infinity, with the property that $R(x_\alpha)$ has a limit, shows that $\lim_{x\to\infty} R(x) = 1$.

Let N denote a (connected component of a) level surface of f. At a point of N, choose an orthonormal frame $\{e_1, e_2, e_3\}$ with $e_3 = X$ normal to N. From the Gauss-Codazzi equation,

$$R^N = 2\,K^N(e_1,e_2) = 2(K^M(e_1,e_2) + \det(S)), \tag{51.13}$$

where S is the shape operator. As $R = 2(K^M(e_1,e_2) + K^M(e_1,e_3) + K^M(e_2,e_3))$ and $\mathrm{Ric}(X,X) = K^M(e_1,e_3) + K^M(e_2,e_3)$, we obtain

$$R^N = R - 2\,\mathrm{Ric}(X,X) + 2\det(S). \tag{51.14}$$

The shape operator is given by $S = \frac{Hess f|_{TN}}{|\nabla f|}$. From (51.2), $\mathrm{Hess}\, f = \frac{1}{2} - \mathrm{Ric}$. We can diagonalize $\mathrm{Ric}\Big|_{TN}$ to write $\mathrm{Ric} = \begin{pmatrix} r_1 & 0 & c_1 \\ 0 & r_2 & c_2 \\ c_1 & c_2 & r_3 \end{pmatrix}$, where $r_3 = \mathrm{Ric}(X,X)$. Then

$$\begin{aligned} \det\left(\mathrm{Hess}\, f\Big|_{TN}\right) &= \left(\frac{1}{2} - r_1\right)\left(\frac{1}{2} - r_2\right) = \frac{1}{4}\left((1-r_1-r_2)^2 - (r_1-r_2)^2\right) \\ &\le \frac{1}{4}(1-r_1-r_2)^2 = \frac{1}{4}(1 - R + \mathrm{Ric}(X,X))^2. \end{aligned} \tag{51.15}$$

This shows that the scalar curvature of N is bounded above by

$$R - 2\,\mathrm{Ric}(X,X) + \frac{(1-R+\mathrm{Ric}(X,X))^2}{2|\nabla f|^2}. \tag{51.16}$$

If $|\nabla f|$ is large then $1 - R + \mathrm{Ric}(X,X) < 2\,|\nabla f|^2$. As $1 - R + \mathrm{Ric}(X,X)$ is positive when the distance from x to x_0 is large enough,

$$\begin{aligned} (1 - R + \mathrm{Ric}(X,X))^2 &< 2(1 - R + \mathrm{Ric}(X,X))|\nabla f|^2 \\ &\le 2(1 - R + \mathrm{Ric}(X,X))|\nabla f|^2 + 2|\nabla f|^2\,\mathrm{Ric}(X,X) \end{aligned} \tag{51.17}$$

and so

$$\frac{(1 - R + \mathrm{Ric}(X,X))^2}{2|\nabla f|^2} < 1 - R + 2\,\mathrm{Ric}(X,X). \tag{51.18}$$

Hence

$$R - 2\,\mathrm{Ric}(X,X) + \frac{(1-R+\mathrm{Ric}(X,X))^2}{2|\nabla f|^2} < 1. \tag{51.19}$$

This shows that $R^N < 1$ if N is sufficiently far from x_0.

If Y is a unit vector that is tangential to N then from (51.2),

$$\nabla_Y \nabla_Y f \; = \; \frac{1}{2} \; - \; \mathrm{Ric}(Y, Y). \tag{51.20}$$

If $\{Y, Z, W\}$ is an orthonormal basis then

$$\mathrm{Ric}(Y, Y) = K^M(Y, Z) + K^M(Y, W) \leq K^M(Y, Z) + K^M(Y, W) + K^M(Z, W) = \frac{1}{2} R. \tag{51.21}$$

Hence $\nabla_Y \nabla_Y f \; \geq \; \frac{1}{2}(1 - R)$, which is positive if N is sufficiently far from x_0. Thus N is convex and so the area of the level set increases as the level increases. On the other hand, we can take points x_α on the level sets going to infinity, apply the previous splitting argument and use the fact that $\operatorname{grad} f$ becomes almost parallel to $\operatorname{grad} d(\cdot, x_0)$. Within one of the approximate cylinders coming from the splitting argument, there is a projection π to its base S^2. As a tangent plane of N is almost perpendicular to $\mathrm{Ker}(d\pi)$, the restriction of π to N is an almost-isometry from N to S^2. By the monotonicity of $\mathrm{area}(N)$, we conclude that $\mathrm{area}(N) \; \leq \; 8\pi$ if N is sufficiently far from x_0. However as N is a topologically a 2-sphere, the Gauss-Bonnet theorem says that $\int_N R^N \, dA \; = \; 8\pi$. This contradicts the facts that $R^N \; < 1$ and $\mathrm{area}(N) \; \leq \; 8\pi$. □

Corollary 51.22. *The only complete oriented 3-dimensional noncompact κ-noncollapsed gradient shrinking solitons with bounded nonnegative sectional curvature are the round evolving $\mathbb{R} \times S^2$ and its $\mathbb{Z}_2$-quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^2$.*

Proof. Let $(M, g(\cdot))$ be a complete oriented 3-dimensional noncompact κ-noncollapsed gradient shrinking soliton with bounded nonnegative sectional curvature. By Lemma 51.1, M cannot have positive sectional curvature. From Theorem A.7, M must locally split off an $\mathbb{R}$-factor. Then the universal cover splits off an $\mathbb{R}$-factor and so, by Corollary 40.1 or Section 43, must be the standard $\mathbb{R} \times S^2$. From the κ-noncollapsing, M must be $\mathbb{R} \times S^2$ or $\mathbb{R} \times_{\mathbb{Z}_2} S^2$. □

52. I.12.1. Canonical neighborhood theorem

In this section we show that a high-curvature region of a three-dimensional Ricci flow is modeled by part of a κ-solution.

We first define the notion of Φ-almost nonnegative curvature.

Definition 52.1. (cf. I.12) Let $\Phi \in C^\infty(\mathbb{R})$ be a positive nondecreasing function such that for positive s, $\frac{\Phi(s)}{s}$ is a decreasing function which tends to zero as $s \to \infty$. A Ricci flow solution is said to have Φ-*almost nonnegative curvature* if for all (x, t), we have

$$\mathrm{Rm}(x, t) \; \geq \; - \, \Phi(R(x, t)). \tag{52.2}$$

Remark 52.3. Note that Φ-almost nonnegative curvature implies that the scalar curvature is uniformly bounded below by $- \, 6 \, \Phi(0)$. The formulation of the pinching condition in [51, Section 12] is that there is a decreasing function ϕ, tending to zero at infinity, so that $\mathrm{Rm}(x, t) \; \geq \; - \, \phi(R(x, t)) \, R(x, t)$ for each (x, t). This formulation has a problem when $R(x, t) < 0$, if one takes ϕ to be defined on all of $\mathbb{R}$. The condition in Definition 52.1 is what

comes out of the three-dimensional Hamilton-Ivey pinching result (applied to the rescaled metric $\tilde{g}(t) = \frac{g(t)}{t}$) if we assume normalized initial conditions; see Appendix B.

We note that since the sectional curvatures have to add up to R, the lower bound (52.2) implies a double-sided bound on the sectional curvatures. Namely,

$$-\Phi(R) \leq \mathrm{Rm} \leq \frac{R}{2} + \left(\frac{n(n-1)}{2} - 1 \right) \Phi(R). \tag{52.4}$$

The main use of the pinching condition is to show that blowup limits have nonnegative sectional curvature.

Lemma 52.5. *Let $\{(M_k, p_k, g_k)\}_{k=1}^{\infty}$ be a sequence of complete pointed Riemannian manifolds with Φ-almost nonnegative curvature. Given a sequence $Q_k \to \infty$, put $\overline{g}_k = Q_k g_k$ and suppose that there is a pointed smooth limit $(M_\infty, p_\infty, \overline{g}_\infty) = \lim_{k \to \infty} (M_k, p_k, \overline{g}_k)$. Then M_∞ has nonnegative sectional curvature.*

Proof. First, the Φ-almost nonnegative curvature condition implies that the scalar curvature of M_k is bounded below uniformly in k. For $m \in M_\infty$, let $m_k \in (M_k, \bar{g}_k)$ be a sequence of approximants to m. Then $\lim_{k \to \infty} \overline{\mathrm{Rm}}(m_k) = \overline{\mathrm{Rm}}_\infty(m)$, where $\overline{\mathrm{Rm}}(m_k) = Q_k^{-1} \mathrm{Rm}(m_k)$. There are two possibilities : either the numbers $R(m_k)$ are uniformly bounded above or they are not. If they are uniformly bounded above then (52.4) implies that $\mathrm{Rm}(m_k)$ is uniformly bounded above and below, so $\overline{\mathrm{Rm}}_\infty(m) = 0$. Suppose on the other hand that a subsequence of the numbers $R(m_k)$ tends to infinity. We pass to this subsequence. Now $\overline{R}_\infty(m) = \lim_{k \to \infty} \overline{R}(m_k)$ exists by assumption and is nonnegative. Applying (52.2) gives that $\overline{\mathrm{Rm}}_\infty(m)$, the limit of

$$Q_k^{-1} \mathrm{Rm}(m_k) = \overline{R}(m_k) \frac{\mathrm{Rm}(m_k)}{R(m_k)}, \tag{52.6}$$

is nonnegative. □

We now prove the first version of the "canonical neighborhood" theorem.

Theorem 52.7. *(cf. Theorem I.12.1) Given $\epsilon, \kappa, \sigma > 0$ and a function Φ as above, one can find $r_0 > 0$ with the following property. Let $g(\cdot)$ be a Ricci flow solution on a three-manifold M, defined for $0 \leq t \leq T$ with $T \geq 1$. We suppose that for each t, $g(t)$ is complete, and the sectional curvature is bounded on compact time intervals. Suppose that the Ricci flow has Φ-almost nonnegative curvature and is κ-noncollapsed on scales less than σ. Then for any point (x_0, t_0) with $t_0 \geq 1$ and $Q = R(x_0, t_0) \geq r_0^{-2}$, the solution in $\{(x,t) : \mathrm{dist}^2_{t_0}(x, x_0) < (\epsilon Q)^{-1}, t_0 - (\epsilon Q)^{-1} \leq t \leq t_0\}$ is, after scaling by the factor Q, ϵ-close to the corresponding subset of a κ-solution.*

Remark 52.8. Our statement of Theorem 52.7 differs slightly from that in [51, Theorem 12.1]. First, we allow M to be noncompact, provided that there is bounded sectional curvature on compact time intervals. This generalization will be useful for later work. More importantly, the statement in [51, Theorem 12.1] has noncollapsing at scales less than r_0, whereas we require noncollapsing at scales less than σ. See Remark 52.19 for further comment.

In the phrase "$t_0 \geq 1$" there is an implied scale which comes from the Φ-almost nonnegativity assumption, and similarly for the statement "scales less than σ".

Proof. We give a proof which differs in some points from the proof in [51] but which has the same ingredients. We first outline the argument.

Suppose that the theorem is false. Then for some $\epsilon, \kappa, \sigma > 0$, we have a sequence of such Φ-nonnegatively curved 3-dimensional Ricci flows $(M_k, g_k(\cdot))$ defined on intervals $[0, T_k]$, and sequences $r_k \to 0$, $\hat{x}_k \in M_k$, $\hat{t}_k \geq 1$ such that M_k is κ-noncollapsed on scales $< \sigma$ and $Q_k = R(\hat{x}_k, \hat{t}_k) \geq r_k^{-2}$, but the Q_k-rescaled solution in $B_{(\epsilon Q_k)^{-1/2}}(\hat{x}_k) \times [\hat{t}_k - (\epsilon Q_k)^{-1}, \hat{t}_k]$ is not ϵ-close to the corresponding subset of any κ-solution.

We note that if the statement were false for ϵ then it would also be false for any smaller ϵ. Because of this, somewhat paradoxically, we will begin the argument with a given ϵ but will allow ourselves to make ϵ small enough later so that the argument works. To be clear, we will eventually get a contradiction using a fixed (small) value of ϵ, but as the proof goes along we will impose some upper bounds on this value in order for the proof to work. (If we tried to list all of the constraints at the beginning of the argument then they would look unmotivated.)

The goal is to get a contradiction based on the "bad" points $(\widehat{x}_k, \widehat{t}_k)$. In a sense, the method of proof of Theorem 52.7 is an induction on the curvature scale. For example, if we were to make the additional assumption in the theorem that $R(x,t) \leq R(x_0, t_0)$ for all $x \in M$ and $t \leq t_0$ then the theorem would be very easy to prove. We would just take a convergent subsequence of the rescaled solutions, based at $(\hat{x}_k, \hat{t}_k)$, to get a κ-solution; this would give a contradiction. This simple argument can be considered to be the first step in a proof by induction on curvature scale. In the proof of Theorem 52.7 one effectively proves the result at a given curvature scale inductively by assuming that the result is true at higher curvature scales.

The actual proof consists of four steps. Step 1 consists of replacing the sequence $(\widehat{x}_k, \widehat{t}_k)$ by another sequence of "bad" points (x_k, t_k) which have the property that points near (x_k, t_k) with distinctly higher scalar curvature are "good" points. It then suffices to get a contradiction based on the existence of the sequence (x_k, t_k).

In steps 2-4 one uses the points (x_k, t_k) to build up a κ-solution, whose existence then contradicts the "badness" of the points (x_k, t_k). More precisely, let $(M_k, (x_k, t_k), \bar{g}_k(\cdot))$ be the result of rescaling $g_k(\cdot)$ by $R(x_k, t_k)$. We will show that the sequence of pointed flows $(M_k, (x_k, t_k), \bar{g}_k(\cdot))$ accumulates on a κ-solution $(M_\infty, (x_\infty, t_0), \overline{g}_\infty(\cdot))$, thereby obtaining a contradiction.

In step 2 one takes a pointed limit of the manifolds $(M_k, x_k, \bar{g}_k(t_k))$ in order to construct what will become the final time slice of the κ-solution, $(M_\infty, x_\infty, \overline{g}_\infty(t_0))$. In order to take this limit, it is necessary to show that the manifolds $(M_k, x_k, \bar{g}_k(t_k))$ have uniformly bounded curvature on distance balls of a fixed radius. If this were not true then for some radius, a subsequence of the manifolds $(M_k, x_k, \bar{g}_k(t_k))$ would have curvatures that asymptotically blowup on the ball of that radius. One shows that geometrically, the curvature blowup is due to the asymptotic formation of a cone-like point at the blowup radius. Doing a further rescaling at this cone-like point, one obtains a Ricci flow solution that ends on a part of a nonflat metric cone. This gives a contradiction as in the case $0 < \mathcal{R} < \infty$ of Theorem 41.2.

Thus one can construct the pointed limit $(M_\infty, x_\infty, \overline{g}_\infty(t_0))$. The goal now is to show that $(M_\infty, x_\infty, \overline{g}_\infty(t_0))$ is the final time slice of a κ-solution $(M_\infty, (x_\infty, t_0), \overline{g}_\infty(\cdot))$. In step 3

one shows that $(M_\infty, x_\infty, \overline{g}_\infty(t_0))$ extends backward to a Ricci flow solution on some time interval $[t_0 - \Delta, t_0]$, and that the time slices have bounded nonnegative curvature. In step 4 one shows that the Ricci flow solution can be extended all the way to time $(-\infty, t_0]$, thereby constructing a κ-solution

Step 1: Adjusting the choice of basepoints.

We first modify the points $(\widehat{x}_k, \widehat{t}_k)$ slightly in time to points (x_k, t_k) so that in the given Ricci flow solution, there are no other "bad" points with much larger scalar curvature in a earlier time interval whose length is large compared to $R(x_k, t_k)^{-1}$. (The phrase "nearly the smallest curvature Q" in [51, Proof of Theorem 12.1] should read "nearly the largest curvature Q". This is clear from the sentence in parentheses that follows.) The proof of the next lemma is by pointpicking, as in Appendix H.

Lemma 52.9. *We can find $H_k \to \infty$, $x_k \in M_k$, and $t_k \geq \frac{1}{2}$ such that $Q_k = R(x_k, t_k) \to \infty$, and for all k the conclusion of Theorem 52.7 fails at (x_k, t_k), but holds for any $(y, t) \in M_k \times [t_k - H_k Q_k^{-1}, t_k]$ for which $R(y, t) \geq 2R(x_k, t_k)$.*

Proof. Choose $H_k \to \infty$ such that $H_k (R(\widehat{x}_k, \widehat{t}_k))^{-1} \leq \frac{1}{10}$ for all k. For each k, initially set $(x_k, t_k) = (\hat{x}_k, \hat{t}_k)$. Put $Q_k = R(x_k, t_k)$ and look for a point in $M_k \times [t_k - H_k Q_k^{-1}, t_k]$ at which Theorem 52.7 fails, and the scalar curvature is at least $2R(x_k, t_k)$. If such a point exists, replace (x_k, t_k) by this point; otherwise do nothing. Repeat this until the second alternative occurs. This process must terminate with a new choice of (x_k, t_k) satisfying the lemma. □

Hereafter we use this modified sequence (x_k, t_k). Let $(M_k, (x_k, t_k), \bar{g}_k(\cdot))$ be the result of rescaling $g_k(\cdot)$ by $Q_k = R(x_k, t_k)$. We use $\bar{R}_k$ to denote its scalar curvature; in particular, $\bar{R}_k(x_k, t_k) = 1$. Note that the rescaled time interval of Lemma 52.9 has duration $H_k \to \infty$; this is what we want in order to try to extract an ancient solution.

Step 2: For every $\rho < \infty$, the scalar curvature $\bar{R}_k$ is uniformly bounded on the ρ-balls $B(x_k, \rho) \subset (M_k, \bar{g}_k(t_k))$ (the argument for this is essentially equivalent to [51, Pf. of Claim 2 of Theorem I.12.1]*).* Before proceeding, we need some bounds which come from our choice of basepoints, and the derivative bounds inherited (by approximation) from κ-solutions.

Lemma 52.10. *There is a constant $C = C(\kappa)$ so that for any (x, t) in a Ricci flow solution, if $R(x, t) > 0$ and the solution in $B_t(x, (\epsilon R(x, t))^{-1/2}) \times [t - (\epsilon R(x, t))^{-1}, t]$ is ϵ-close to a corresponding subset of a κ-solution then $|\nabla R^{-1/2}|(x, t) \leq C$ and $|\partial_t R^{-1}|(x, t) \leq C$.*

Proof. This follows from the compactness in Theorem 46.1. □

Note that the same value of C in Lemma 52.10 also works for smaller ϵ.

Lemma 52.11. *(cf. Claim 1 of I.12.1) For each $(\overline{x}, \overline{t})$ with $t_k - \frac{1}{2} H_k Q_k^{-1} \leq \overline{t} \leq t_k$, we have $R_k(x, t) \leq 4\overline{Q}_k$ whenever $\overline{t} - c\, \overline{Q}_k^{-1} \leq t \leq \overline{t}$ and $\mathrm{dist}_{\overline{t}}(x, \overline{x}) \leq c\, \overline{Q}_k^{-1/2}$, where $\overline{Q}_k = Q_k + |R_k(\overline{x}, \overline{t})|$ and $c = c(\kappa) > 0$ is a small constant.*

Proof. If $R_k(x,t) \le 2Q_k$ then there is nothing to show. If $R_k(x,t) > 2Q_k$, consider a spacetime curve γ that goes linearly from (x,t) to $(x,\overline{t})$, and then goes from $(x,\overline{t})$ to $(\overline{x},\overline{t})$ along a minimizing geodesic. If there is a point on γ with curvature $2Q_k$, let p be the nearest such point to (x,t). If not, put $p = (\overline{x},\overline{t})$. From the conclusion of Lemma 52.9, we can apply Lemma 52.10 along γ from (x,t) to p. The claim follows from integrating the ensuing derivative bounds along γ. □

Lemma 52.12. *In terms of the rescaled solution $\overline{g}_k(\cdot)$, for each $(\overline{x},\overline{t})$ with $t_k - \frac{1}{2}H_k \le \overline{t} \le t_k$, we have $\overline{R}_k(x,t) \le 4\widetilde{Q}_k$ whenever $\overline{t} - c\,\widetilde{Q}_k^{-1} \le t \le \overline{t}$ and $\mathrm{dist}_{\overline{t}}(x,\overline{x}) \le c\,\widetilde{Q}_k^{-1/2}$, where $\widetilde{Q}_k = 1 + |\overline{R}_k(\overline{x},\overline{t})|$.*

Proof. This is just the rescaled version of Lemma 52.11. □

For all $\rho \ge 0$, put

$$D(\rho) = \sup\{\bar{R}_k(x,t_k) \mid k \ge 1,\ x \in B(x_k,\rho) \subset (M_k, \bar{g}_k(t_k))\}, \tag{52.13}$$

and let ρ_0 be the supremum of the ρ's for which $D(\rho) < \infty$. Note that $\rho_0 > 0$, in view of Lemma 52.12 (taking $(\bar{x},\bar{t}) = (x_k,t_k)$). Suppose that $\rho_0 < \infty$. After passing to a subsequence if necessary, we can find a sequence $y_k \in M_k$ with $\mathrm{dist}_{t_k}(x_k,y_k) \to \rho_0$ and $\bar{R}(y_k,t_k) \to \infty$. Let $\eta_k \subset (M_k,\bar{g}_k(t_k))$ be a minimizing geodesic segment from x_k to y_k. Let $z_k \in \eta_k$ be the point on η_k closest to y_k at which $\bar{R}(z_k,t_k) = 2$, and let γ_k be the subsegment of η_k running from y_k to z_k. By Lemma 52.12 the length of γ_k is bounded away from zero independent of k. Due to the Φ-pinching (see (52.4)), for all $\rho < \rho_0$, we have a uniform bound on $|\operatorname{Rm}|$ on the balls $B(x_k,\rho) \subset (M_k,\bar{g}_k(t_k))$. The injectivity radius is also controlled in $B(x_k,\rho)$, in view of the curvature bounds and the κ-noncollapsing. Therefore after passing to a subsequence, we can assume that the pointed sequence $(B(x_k,\rho_0),\bar{g}_k(t_k),x_k)$ converges in the pointed Gromov-Hausdorff topology (i.e. for all $\rho < \rho_0$ we have the usual Gromov-Hausdorff convergence) to a pointed C^1-Riemannian manifold $(Z,\bar{g}_\infty,x_\infty)$, the segments η_k converge to a segment (missing an endpoint) $\eta_\infty \subset Z$ emanating from x_∞, and γ_k converges to $\gamma_\infty \subset \eta_\infty$. Let $\bar{Z}$ denote the completion of $(Z,\bar{g}_\infty)$, and $y_\infty \in \bar{Z}$ the limit point of η_∞. Note that by Lemma 52.9 and part 4 of Corollary 42.1, the Riemannian structure near γ_∞ may be chosen to be many times differentiable. (Alternatively, this follows from Lemma 52.12 and the Shi estimates of Appendix D.) In particular the scalar curvature $\bar{R}_\infty$ is defined, differentiable, and satisfies the bound in Lemma 52.10 near γ_∞.

Lemma 52.14. *1. There is a function $c : (0,\infty) \to \mathbb{R}$ depending only on κ, with $\lim_{t\to 0} c(t) = \infty$, such that if $w \in \gamma_\infty$ then $\bar{R}_\infty(w)\,d(y_\infty,w)^2 > c(\epsilon)$.*

2. There is a function $\epsilon' : (0,\infty) \to \mathbb{R} \cup \{\infty\}$ depending only on κ, with $\lim_{t\to 0}\epsilon'(t) = 0$, such that if $w \in \gamma_\infty$ and $d(y_\infty,w)$ is sufficiently small then the pointed manifold $(Z,w,\bar{R}_\infty(w)\bar{g}_\infty)$ is $2\epsilon'(\epsilon)$-close to a round cylinder in the C^2 topology.

Proof. It follows from Lemma 52.9 that for all $w \in \gamma_\infty$, the pointed Riemannian manifold $(Z,w,\bar{R}_\infty(w)\bar{g}_\infty)$ is 2ϵ-close to (a time slice of) a pointed κ-solution. From the definition of pointed closeness, there is an embedded region around w, large on the scale defined by $\bar{R}_\infty(w)$, which is close to the corresponding subset of a pointed κ-solution. This gives a lower bound on the distance $\rho_0 - d(w,x_\infty)$ to the point of curvature blowup, thereby proving part 1 of the lemma.

We know that $(Z, w, \bar{R}_\infty(w)\bar{g}_\infty)$ is 2ϵ-close to a pointed κ-solution $(N, \star, h(t))$ in the pointed C^2-topology. By Lemma 52.10, we know $\bar{R}_\infty(w)$ tends to ∞ as $d(w, x_\infty) \to \rho_0$, for $w \in \gamma_\infty$. In particular, $\bar{R}_\infty(w) d^2(w, x_\infty) \to \infty$. From part 1 above, we can choose ϵ small enough in order to make $\bar{R}_\infty(w) d^2(w, y_\infty)$ large enough to apply Proposition 49.1. Hence the pointed manifold $(N, \star, R(\star)h(t))$ is $\epsilon''(\epsilon)$-close to a round cylinder, where $\epsilon'' : (0, \infty) \to \mathbb{R}$ is a function with $\lim_{t\to 0} \epsilon''(t) = 0$. The lemma follows. $\square$

Note that the function ϵ' in Lemma 52.14 is independent of the particular manifold Z that arises in our proof.

From part 2 of Lemma 52.14, if ϵ is small and $w \in \gamma_\infty$ is sufficiently close to y_∞ then $(Z, w, \overline{R}_\infty(w)\overline{g}_\infty)$ is C^2-close to a round cylinder. The cross-section of the cylinder has diameter approximately $\pi(\overline{R}_\infty(w)/2)^{-\frac{1}{2}}$. If we form the union of the balls $B(w, 2\pi(R_\infty(w)/2)^{-\frac{1}{2}})$, as w ranges over such points in γ_∞, then we obtain a connected Riemannian manifold W. By adding in the point y_∞, we get a metric space $\bar{W}$ which is locally complete, and geodesic near y_∞. As the original manifolds M_k had Φ-almost nonnegative curvature, it follows from Lemma 52.5 that W is nonnegatively curved. Furthermore, y_∞ cannot be an interior point of any geodesic segment in $\bar{W}$, since such a geodesic would have to pass through a cylindrical region near y_∞ twice. The usual proof of the Toponogov triangle comparison inequality now applies near y_∞ since minimizers remain in the smooth nonnegatively curved part of $\bar{W}$. Then $\overline{W}$ has nonnegative curvature in the Alexandrov sense.

This implies that blowups of $(\bar{W}, y_\infty)$ converge to the tangent cone $C_{y_\infty}\bar{W}$. As $\overline{W}$ is three-dimensional, so is $C_{y_\infty}\bar{W}$ [12, Corollary 7.11]. It will be C^2-smooth away from the vertex and nowhere flat, by part 2 of Lemma 52.14. Pick $z \in C_{y_\infty}\bar{W}$ such that $d(z, y_\infty) = 1$. Then the ball $B(z, \frac{1}{2}) \subset C_{y_\infty}\bar{W}$ is the Gromov-Hausdorff limit of a sequence of rescaled balls $B(\widehat{z}_k, \widehat{r}_k) \subset (M_k, \bar{g}_k(t_k))$ where $\widehat{r}_k \to 0$, whose center points $(\widehat{z}_k, t_k)$ satisfy the conclusions of Theorem 52.7. Applying Lemma 52.12 and Appendix B, we get the curvature bounds needed to extract a limiting Ricci flow solution whose time zero slice is isometric to $B(z, \frac{1}{2})$. Now we can apply the reasoning from the $0 < \mathcal{R} < \infty$ case of Theorem 41.2 to get a contradiction. This completes step 2.

Step 3: The sequence of pointed flows $(M_k, \bar{g}_k(\cdot), (x_k, t_k))$ accumulates on a pointed Ricci flow $(M_\infty, \bar{g}_\infty(\cdot), (x_\infty, t_0))$ which is defined on a time interval $[t', t_0]$ with $t' < t_0$. By step 2, we know that the scalar curvature of $(M_k, \bar{g}_k(t_k))$ at $y \in M_k$ is bounded by a function of the distance from y to x_k. Lemma 52.12 extends this curvature control to a backward parabolic neighborhood centered at y whose radius depends only on $d(y, x_k)$. Thus we can conclude, using Φ-pinching (52.4) and Shi's estimates (Appendix D), that all derivatives of the curvature $(M_k, \bar{g}_k(t_k))$ are controlled as a function of the distance from x_k, which means that the sequence of pointed manifolds $(M_k, \bar{g}_k(t_k), x_k)$ accumulates to a smooth manifold $(M_\infty, \bar{g}_\infty)$.

From Lemma 52.5, M_∞ has nonnegative sectional curvature. We claim that M_∞ has bounded curvature. If not then there is a sequence of points $q_k \in M_\infty$ so that $\lim_{k\to\infty} \overline{R}(q_k) = \infty$ and $\overline{R}(q) \le 2\overline{R}(q_k)$ for $q \in B(y_k, A_k \overline{R}(q_k)^{-\frac{1}{2}})$, where $A_k \to \infty$; compare [33, Lemma 22.2]. Lemma 52.9 implies that for large k, a rescaled neighborhood of (M_∞, q_k) is ϵ-close to the corresponding subset of a time slice of a κ-solution. As in the proof of Theorem 46.1, we

obtain a sequence of neck-like regions in M_∞ with smaller and smaller cross-sections, which contradicts the existence of the Sharafutdinov retraction.

By Lemma 52.12 again, we now get curvature control on $(M_k, \bar{g}_k(\cdot))$ for a time interval $[t_k - \Delta, t_k]$ for some $\Delta > 0$, and hence we can extract a subsequence which converges to a pointed Ricci flow $(M_\infty, (x_\infty, t_0), \bar{g}_\infty(\cdot))$ defined for $t \in [t_0 - \Delta, t_0]$, which has nonnegative curvature and bounded curvature on compact time intervals.

Step 4: Getting an ancient solution. Let $(t', t_0]$ be the maximal time interval on which we can extract a limiting solution $(M_\infty, \bar{g}_\infty(\cdot))$ with bounded curvature on compact time intervals. Suppose that $t' > -\infty$. By Lemma 52.12 the maximum of the scalar curvature on the time slice $(M_\infty, \bar{g}_\infty(t))$ must tend to infinity as $t \to t'$. From the trace Harnack inequality, $R_t + \frac{R}{t-t'} \geq 0$, and so

$$\bar{R}_\infty(x,t) \leq Q\frac{t_0 - t'}{t - t'}, \tag{52.15}$$

where Q is the maximum of the scalar curvature on $(M_\infty, \bar{g}_\infty(t_0))$. Combining this with Corollary 27.16, we get

$$\frac{d}{dt} d_t(x,y) \geq \text{const.} \sqrt{Q\frac{t_0 - t'}{t - t'}}. \tag{52.16}$$

Since the right hand side is integrable on $(t', t_0]$, and using the fact that distances are nonincreasing in time (since $\mathrm{Rm} \geq 0$), it follows that there is a constant C such that

$$|d_t(x,y) - d_{t_0}(x,y)| < C \tag{52.17}$$

for all $x, y \in M_\infty$, $t \in (t', t_0]$.

If M_∞ is compact then by (52.17) the diameter of $(M_\infty, \bar{g}_\infty(t))$ is bounded independent of $t \in (t', t_0]$. Since the minimum of the scalar curvature is increasing in time, it is also bounded independent of t. Now the argument in Step 2 shows that the curvature is bounded everywhere independent of t. (We can apply the argument of Step 2 to the time-t slice because the main ingredient was Lemma 52.9, which holds for rescaled time t.)

We may therefore assume M_∞ is noncompact. To be consistent with the notation of I.12.1, we now relabel the basepoint (x_∞, t_0) as (x_0, t_0). Since nonnegatively curved manifolds are asymptotically conical (see Appendix G), there is a constant D such that if $y \in M_\infty$, and $d_{t_0}(y, x_0) > D$, then there is a point $x \in M_\infty$ such that

$$d_{t_0}(x,y) = d_{t_0}(y,x_0) \quad \text{and} \quad d_{t_0}(x,x_0) \geq \frac{3}{2} d_{t_0}(y,x_0); \tag{52.18}$$

by (52.17) the same conditions hold at all times $t \in (t', t_0]$, up to error C. If for some such y, and some $t \in (t', t_0]$ the scalar curvature were large, then $(M_\infty, (y,t), \bar{R}_\infty(y,t)\bar{g}_\infty(t))$ would be 2ϵ-close to a κ-solution $(N, h(\cdot), (z, t_0))$. When ϵ is small we could use Proposition 49.1 to see that y lies in a neck region U in $(M_\infty, \bar{g}_\infty(t))$ of diameter $\approx R(y,t)^{-\frac{1}{2}} \ll 1$.

We claim that U separates x_0 from x in the sense that x_0 and x belong to disjoint components of $M_\infty - U$, where x_0, y, and x satisfy (52.18). To see this, we recall that if M_∞ has more than one end then it splits isometrically, in which case the claim is clear. If M_∞ has one end then we consider an exhaustion of M_∞ by totally convex compact sets C_t as

in Section 46. One of the sets C_t will have boundary consisting of an approximate 2-sphere cross-section in the neck region U, giving the separation.

(For another argument, suppose that U does not separate x_0 from x. Since $\widetilde{\angle}_y(x_0, x) > \frac{\pi}{2}$ (from Proposition 49.1), the segments $\overline{yx_0}$ and $\overline{yx}$ must exit U by different ends. If x_0 and x can be joined by a curve avoiding U then there is a nonzero element of $\mathrm{H}^1(M_\infty, \mathbb{Z})$. The corresponding infinite cyclic cover of M_∞ will then isometrically split off a line and its quotient M_∞ will be compact, which is a contradiction. We thank one of the referees for this argument.)

Obviously, at time t_0 the set U still separates x_0 from x. Since $\bar{g}_\infty$ has nonnegative curvature, we have $\mathrm{diam}_{t_0}(U) \le \mathrm{diam}_t(U) \ll 1$. Since $(M_\infty, \bar{g}_\infty(t_0))$ has bounded geometry, there cannot be topologically separating subsets of arbitrarily small diameter. Thus there must be a uniform upper bound on $R(y,t)$ and the curvature of $(M_\infty, \bar{g}_\infty)$ is uniformly bounded (in space and time) outside a set of uniformly bounded diameter. Repeating the reasoning from Step 2, we get uniform bounds everywhere. This contradicts our assumption that the curvature blows up as $t \to t'$.

It remains to show that the ancient solution is a κ-solution. The only remaining point is to show that it is κ-noncollapsed at all scales. This follows from the fact that the original Ricci flow solutions $(M_k, g_k(\cdot))$ were κ-noncollapsed on scales less than the fixed number σ. □

Remark 52.19. As mentioned in Remark 52.8, the statement of [51, Theorem 12.1] instead assumes noncollapsing at all scales less than r_0. Bing Wang pointed out that with this assumption, after constructing the ancient solution in Step 4 of the proof, one only gets that it is κ-collapsed at all scales less than one. Hence it may not be a κ-solution. The literal statement of [51, Theorem 12.1] is not used in the rest of [51, 52], but rather its method of proof. Because of this, the change of hypotheses does not seem to lead to any problems. The method of proof of Theorem 52.7 is used in two different ways. The first way is to construct the Ricci flow with surgery on a fixed finite time interval, as in Section 77. In this case the noncollapsing at a given scale σ comes from Theorem 26.2, and its extension when surgeries are allowed. The second way is to analyze the large-time behavior of the Ricci flow, as in the next few sections.

53. I.12.2. Later scalar curvature bounds on bigger balls from curvature and volume bounds

The next theorem roughly says that if one has a sectional curvature bound on a ball, for a certain time interval, and a lower bound on the volume of the ball at the initial time, then one obtains an upper scalar curvature bound on a larger ball at the final time.

We first write out the corrected version of the theorem (see II.6.2).

Theorem 53.1. *For any $A < \infty$, there exist $K = K(A) < \infty$ and $\rho = \rho(A) > 0$ with the following property. Suppose in dimension three we have a Ricci flow solution with Φ-almost nonnegative curvature. Given $x_0 \in M$ and $r_0 > 0$, suppose that $r_0^2\, \Phi(r_0^{-2}) < \rho$, the solution is defined for $0 \le t \le r_0^2$ and it has $|\mathrm{Rm}|(x,t) \le \frac{1}{3r_0^2}$ for all (x,t) satisfying*

$\mathrm{dist}_0(x, x_0) < r_0$. *Suppose in addition that the volume of the metric ball $B(x_0, r_0)$ at time zero is at least $A^{-1} r_0^3$. Then $R(x, r_0^2) \le K r_0^{-2}$ whenever $\mathrm{dist}_{r_0^2}(x, x_0) < A r_0$.*

Remark 53.2. The added restriction that $r_0^2 \Phi(r_0^{-2}) < \rho$ (see II.6.2) imposes an upper bound on r_0. This is necessary, as otherwise the conclusion would imply that neck pinches cannot occur.

There is an apparent gap in the proof of [51, Theorem 12.2], in the sentence (There is a little subtlety...). We instead follow the proof of II.6.3(b,c) (see Proposition 84.1(b,c)), which proves the same statement in the presence of surgeries.

The volume assumption in the theorem is used to guarantee noncollapsing, by means of Theorem 28.2. The reason for the "3" in the hypothesis $|\mathrm{Rm}|(x,t) \le \frac{1}{3r_0^2}$ comes from Remark 28.3.

Proof. The proof is in two steps. In the first step one shows that if $R(x, r_0^2)$ is large then a parabolic neighborhood of (x, r_0^2) is close to the corresponding subset of a κ-solution. In the second part one uses this to prove the theorem.

The first step is the following lemma.

Lemma 53.3. *For any $\epsilon > 0$ there exists $K = K(A, \epsilon) < \infty$ so that for any r_0, whenever we have a solution as in the statement of the theorem and $\mathrm{dist}_{r_0^2}(x, x_0) < A r_0$ then*
(a) $R(x, r_0^2) < K r_0^{-2}$ or
(b) The solution in $\{(x', t') : \mathrm{dist}_t(x', x) < (\epsilon Q)^{-1}, t - (\epsilon Q)^{-1} \le t' \le t\}$ is, after scaling by the factor Q, ϵ-close to the corresponding subset of a κ-solution.

Here $t = r_0^2$ and $Q = R(x,t)$.

Remark 53.4. One can think of this lemma as a localized analog of Theorem 52.7, where "localized" refers to the fact that both the hypotheses and the conclusion involve the point x_0.

Proof. To prove the lemma, suppose that there is a sequence of such pointed solutions $(M_k, x_{0,k}, g_k(\cdot))$, along with points $\hat{x}_k \in M_k$, so that $\mathrm{dist}_{r_0^2}(\hat{x}_k, x_{0,k}) < A r_0$ and $r_0^2 R(\hat{x}_k, r_0^2) \to \infty$, but $(\hat{x}_k, r_0^2)$ does not satisfy conclusion (b) of the lemma. As in the proof of Theorem 52.7, we will allow ourselves to make ϵ smaller during the course of the proof.

We first show that there is a sequence $D_k \to \infty$ and modified points $(\overline{x}_k, \overline{t}_k)$ with $\frac{3}{4} r_0^2 \le \overline{t}_k \le r_0^2$, $\mathrm{dist}_{\overline{t}_k}(\overline{x}_k, x_{0,k}) < (A+1) r_0$ and $Q_k = R(\overline{x}_k, \overline{t}_k) \to \infty$, so that any point (x'_k, t'_k) with $R(x'_k, t'_k) > 2Q_k$, $\overline{t}_k - D_k^2 Q_k^{-1} \le t'_k \le \overline{t}_k$ and $\mathrm{dist}_{t'_k}(x'_k, x_{0,k}) < \mathrm{dist}_{\overline{t}_k}(\overline{x}_k, x_{0,k}) + D_k Q_k^{-1/2}$ satisfies conclusion (b) of the lemma, but $(\overline{x}_k, \overline{t}_k)$ does not satisfy conclusion (b) of the lemma. (Of course, in saying "(x'_k, t'_k) satisfies conclusion (b)" or "$(\overline{x}_k, \overline{t}_k)$ does not satisfy conclusion (b)", we mean that the (x,t) in conclusion (b) is replaced by (x'_k, t'_k) or $(\overline{x}_k, \overline{t}_k)$, respectively.)

The construction of $(\overline{x}_k, \overline{t}_k)$ is by a pointpicking argument. Put $D_k = \frac{r_0 R(\hat{x}_k, r_0^2)^{1/2}}{10}$. Start with $(x_k, t_k) = (\hat{x}_k, r_0^2)$ and look if there is a point (x'_k, t'_k) with $R(x'_k, t'_k) > 2R(x_k, t_k)$, $\overline{t}_k - D_k^2 R(x_k, t_k)^{-1} \le t'_k \le \overline{t}_k$ and $\mathrm{dist}_{t'_k}(x'_k, x_{0,k}) < \mathrm{dist}_{t_k}(x_k, x_{0,k}) + D_k R(x_k, t_k)^{-1/2}$, but

which does not have a neighborhood that is ϵ-close to the corresponding subset of a κ-solution. If there is such a point, we replace (x_k, t_k) by (x'_k, t'_k) and repeat the process. The process must terminate after a finite number of steps to give a point $(\overline{x}_k, \overline{t}_k)$ with the desired property.

(Note that the condition $\mathrm{dist}_{t'_k}(x'_k, x_{0,k}) < \mathrm{dist}_{\overline{t}_k}(\overline{x}_k, x_{0,k}) + D_k Q_k^{-1/2}$ involves the metric at time t'_k. In order to construct an ancient solution, one of the issues will be to replace this by a condition that only involves the metric at time $\overline{t}_k$, i.e. that involves a parabolic neighborhood around $(\overline{x}_k, \overline{t}_k)$.)

Let $\overline{g}_k(\cdot)$ denote the rescaling of the solution $g_k(\cdot)$ by Q_k. We normalize the time interval of the rescaled solution by fixing a number t_∞ and saying that for all k, the time-$\overline{t}_k$ slice of (M_k, g_k) corresponds to the time-t_∞ slice of $(M_k, \overline{g}_k)$. Then the scalar curvature $\overline{R}_k$ of $\overline{g}_k$ satisfies $\overline{R}_k(\overline{x}_k, t_\infty) = 1$.

By the argument of Step 2 of the proof of Theorem 52.7, a subsequence of the pointed spaces $(M_k, \overline{x}_k, \overline{g}_k(t_\infty))$ will smoothly converge to a nonnegatively-curved pointed space $(M_\infty, \overline{x}_\infty, \overline{g}_\infty)$. By the pointpicking, if $m \in M_\infty$ has $\overline{R}(m) \geq 3$ then a parabolic neighborhood of m is ϵ-close to the corresponding region in a κ-solution. It follows, as in Step 3 of the proof of Theorem 52.7, that the sectional curvature of M_∞ will be bounded above by some $C < \infty$. Using Lemma 52.12, the metric on M_∞ is the time-t_∞ slice of a nonnegatively-curved Ricci flow solution defined on some time interval $[t_\infty - c, t_\infty]$, with $c > 0$, and one has convergence of a subsequence $\overline{g}_k(t) \rightarrow \overline{g}_\infty(t)$ for $t \in [t_\infty - c, t_\infty]$. As $\overline{R}_t \geq 0$, the scalar curvature on this time interval will be uniformly bounded above by $6C$ and so from the Φ-almost nonnegative curvature (see (52.4)), the sectional curvature will be uniformly bounded above on the time interval. Hence we can apply Lemma 27.8 to get a uniform additive bound on the length distortion between times $t_\infty - c$ and t_∞ (see Step 4 of the proof of Theorem 52.7). More precisely, in applying Lemma 27.8, we use the curvature bound coming from the hypotheses of the theorem near x_0, and the just-derived upper curvature bound near $\overline{x}_k$.

It follows that for a given $A' > 0$, for large k, if $t'_k \in [\overline{t}_k - cQ_k^{-1}/2, \overline{t}_k]$ and $\mathrm{dist}_{\overline{t}_k}(x'_k, \overline{x}_k) < A' Q_k^{-1/2}$ then $\mathrm{dist}_{t'_k}(x'_k, x_{0,k}) < \mathrm{dist}_{\overline{t}_k}(\overline{x}_k, x_{0,k}) + D_k Q_k^{-1/2}$. In particular, if a point (x'_k, t'_k) lies in the parabolic neighborhood given by $t'_k \in [\overline{t}_k - cQ_k^{-1}/2, \overline{t}_k]$ and $\mathrm{dist}_{\overline{t}_k}(x'_k, \overline{x}_k) < A' Q_k^{-1/2}$, and has $R(x'_k, t'_k) > 2Q_k$, then it has a neighborhood that is ϵ-close to the corresponding subset of a κ-solution.

As in Step 4 of the proof of Theorem 52.7, we now extend $(M_\infty, \overline{g}_\infty, \overline{x}_\infty)$ backward to an ancient solution $\overline{g}_\infty(\cdot)$, defined for $t \in (-\infty, t_\infty]$. To do so, we use the fact that if the solution is defined backward to a time-t slice then the length distortion bound, along with the pointpicking, implies that a point m in a time-t slice with $\overline{R}_\infty(m) > 3$ has a neighborhood that is ϵ-close to the corresponding subset of a κ-solution. The ancient solution is κ-noncollapsed at all scales since the original solution was κ-noncollapsed at some scale, by Theorem 28.2. Then we obtain smooth convergence of parabolic regions of the points $(\overline{x}_k, \overline{t}_k)$ to the κ-solution, which is a contradiction to the choice of the $(\overline{x}_k, \overline{t}_k)$'s. $\square$

We now know that regions of high scalar curvature are modeled by corresponding regions in κ-solutions. To continue with the proof of the theorem, fix A large. Suppose that the

theorem is not true. Then there are
1. Numbers $\rho_k \to 0$,
2. Numbers $r_{0,k}$ with $r_{0,k}^2 \, \Phi(r_{0,k}^{-2}) \,\leq\, \rho_k$,
3. Solutions $(M_k, g_k(\cdot))$ defined for $0 \leq t \leq r_{0,k}^2$,
4. Points $x_{0,k} \in M_k$ and
5. Points $x_k \in M_k$
so that
a. $|\operatorname{Rm}|(x,t) \leq r_{0,k}^{-2}$ for all $(x,t) \in M_k \times [0, r_{0,k}^2]$ satisfying $\operatorname{dist}_0(x, x_{0,k}) < r_{0,k}$,
b. The volume of the metric ball $B(x_{0,k}, r_{0,k})$ at time zero is at least $A^{-1} r_{0,k}^3$ and
c. $\operatorname{dist}_{r_{0,k}^2}(x_k, x_{0,k}) < A r_{0,k}$, but
d. $r_{0,k}^2 \, R(x_k, r_{0,k}^2) \to \infty$.

We now apply Step 2 of the proof of Theorem 52.7 to obtain a contradiction. That is, we take a subsequence of $\{(M_k, x_{0,k}, r_{0,k}^{-2} \, g_k(r_{0,k}^2))\}_{k=1}^{\infty}$ that converges on a maximal ball. The only difference is that in Theorem 52.7, the nonnegative curvature of W came from the Φ-almost nonnegative curvature assumption on the original manifolds M_k along with the fact (with the notation of the proof of Theorem 52.7) that the numbers $Q_k = R(x_k, t_k)$, which we used to rescale, go to infinity. In the present case the rescaled scalar curvatures $r_{0,k}^2 \, R(x_{0,k}, r_{0,k}^2)$ at the basepoints $x_{0,k}$ stay bounded. However, if a point $y \in W$ is a limit of points $\widetilde{x}_k \in M_k$ then the equations

$$\operatorname{Rm}(\widetilde{x}_k, r_{0,k}^2) \;\geq\; -\,\Phi(R(\widetilde{x}_k, r_{0,k}^2)) \tag{53.5}$$

in the form

$$r_{0,k}^2 \, \operatorname{Rm}(\widetilde{x}_k, r_{0,k}^2) \;\geq\; -\,\frac{\Phi\left(r_{0,k}^2 R(\widetilde{x}_k, r_{0,k}^2) \cdot r_{0,k}^{-2}\right)}{R(\widetilde{x}_k, r_{0,k}^2)} \, r_{0,k}^2 \, R(\widetilde{x}_k, r_{0,k}^2) \tag{53.6}$$

pass to the limit to give $\operatorname{Rm}(y) \geq 0$ (using that $y \in W$, so $r_{0,k}^2 \, \operatorname{Rm}(\widetilde{x}_k, r_{0,k}^2) \to R(y) > 0$). This is enough to carry out the argument. $\square$

54. I.12.3. Earlier scalar curvature bounds on smaller balls from lower curvature bounds and volume bounds

The main result of this section says that if one has a lower bound on volume and sectional curvature on a ball at a certain time then one obtains an upper scalar curvature bound on a smaller ball at an earlier time.

We first prove a result in Riemannian geometry saying that under certain hypotheses, metric balls have subballs of a controlled size with almost-Euclidean volume.

Lemma 54.1. *Given $w' > 0$ and $n \in \mathbb{Z}^+$, there is a number $c = c(w', n) > 0$ with the following property. Let B be a radius-r ball with compact closure in an n-dimensional Riemannian manifold. Suppose that the sectional curvatures of B are bounded below by $-r^{-2}$. Suppose that $\operatorname{vol}(B) \geq w' r^n$. Then there is a subball $B' \subset B$ of radius $r' \geq cr$ so that $\operatorname{vol}(B') \,\geq\, \frac{1}{2}\, \omega_n \, (r')^n$, where ω_n is the volume of the unit ball in $\mathbb{R}^n$.*

Proof. Suppose that the lemma is not true. Rescale so that $r = 1$. Then there is a sequence of Riemannian manifolds $\{M_i\}_{i=1}^{\infty}$ with balls $B(x_i, 1) \subset M_i$ having compact closure

so that $\mathrm{Rm}\Big|_{B(x_i,1)} \geq -1$ and $\mathrm{vol}(B(x_i,1)) \geq w'$, but with the property that all balls $B(x_i',r') \subset B(x_i,1)$ with $r' \geq i^{-1}$ satisfy $\mathrm{vol}(B(x_i',r')) < \frac{1}{2}\,\omega_n(r')^n$. After taking a subsequence, we can assume that $\lim_{i\to\infty}(B(x_i,1),x_i) = (X,x_\infty)$ in the pointed Gromov-Hausdorff topology. From [12, Theorem 10.8], the Riemannian volume forms dvol_{M_i} converge weakly to the three-dimensional Hausdorff measure μ of X. From [12, Corollary 6.7 and Section 9], for any $\epsilon > 0$, there are small balls $B(x_\infty',r') \subset X$ with compact closure in X such that $\mu(B(x_\infty',r')) \geq (1-\epsilon)\,\omega_n(r')^n$. This gives a contradiction. □

Theorem 54.2. *(cf. Theorem I.12.3) For any $w > 0$ there exist $\tau = \tau(w) > 0$, $K = K(w) < \infty$ and $\rho = \rho(w) > 0$ with the following property. Suppose that $g(\cdot)$ is a Ricci flow on a closed three-manifold M, defined for $t \in [0,T)$, with Φ-almost nonnegative curvature. Let (x_0,t_0) be a spacetime point and let $r_0 > 0$ be a radius with $t_0 \geq 4\tau r_0^2$ and $r_0^2\,\Phi(r_0^{-2}) < \rho$. Suppose that $\mathrm{vol}_{t_0}(B(x_0,r_0)) \geq w r_0^3$ and the time-t_0 sectional curvatures on $B(x_0,r_0)$ are bounded below by $-\,r_0^{-2}$. Then $R(x,t) \leq K r_0^{-2}$ whenever $t \in [t_0-\tau r_0^2, t_0]$ and $\mathrm{dist}_t(x,x_0) \leq \frac{1}{4}\,r_0$.*

Proof. Let $\tau_0(w)$, $B(w)$ and $C(w)$ be the constants of Corollary 45.13. Put $\tau(w) = \frac{1}{2}\,\tau_0(w)$ and $K(w) = C(w) + 2\frac{B(w)}{\tau_0(w)}$. The function $\rho(w)$ will be specified in the course of the proof.

Suppose that the theorem is not true. Take a counterexample with a point (x_0,t_0) and a radius $r_0 > 0$ such that the time-t_0 ball $B(x_0,r_0)$ satisfies the assumptions of the theorem, but the conclusion of the theorem fails. We claim there is a counterexample coming from a point $(\widehat{x}_0,\widehat{t}_0)$ and a radius $\widehat{r}_0 > 0$, with the additional property that for any (x_0',t_0') and r_0' having $t_0' \in [\widehat{t}_0 - 2\tau\widehat{r}_0^2, \widehat{t}_0]$ and $r_0' \leq \frac{1}{2}\widehat{r}_0$, if $\mathrm{vol}_{t_0'}(B(x_0',r_0')) \geq w(r_0')^3$ and the time-t_0' sectional curvatures on $B(x_0',r_0')$ are bounded below by $-\,(r_0')^{-2}$ then $R(x,t) \leq K(r_0')^{-2}$ whenever $t \in [t_0' - \tau(r_0')^2, t_0']$ and $\mathrm{dist}_t(x,x_0') \leq \frac{1}{4}\,r_0'$. This follows from a pointpicking argument - suppose that it is not true for the original x_0, t_0, r_0. Then there are (x_0',t_0') and r_0' with $t_0' \in [t_0 - 2\tau r_0^2, t_0]$ and $r_0' \leq \frac{1}{2}r_0$, for which the assumptions of the theorem hold but the conclusion does not. If the triple (x_0',t_0',r_0') satisfies the claim then we stop, and otherwise we iterate the procedure. The iteration must terminate, which provides the desired triple $(\widehat{x}_0,\widehat{t}_0,\widehat{r}_0)$. Note that $\widehat{t}_0 > t_0 - 4\tau r_0^2 \geq 0$.

We relabel $(\widehat{x}_0,\widehat{t}_0,\widehat{r}_0)$ as (x_0,t_0,r_0). For simplicity, let us assume that the time-t_0 sectional curvatures on $B(x_0,r_0)$ are strictly greater than $-\,r_0^{-2}$; the general case will follow from continuity. Let $\tau' > 0$ be the largest number such that $\mathrm{Rm}(x,t) \geq -\,r_0^{-2}$ whenever $t \in [t_0 - \tau' r_0^2, t_0]$ and $\mathrm{dist}_t(x,x_0) \leq r_0$. If $\tau' \geq 2\tau = \tau_0(w)$ then Corollary 45.13 implies that $R(x,t) \leq C r_0^{-2} + B(t - t_0 + 2\tau r_0^2)^{-1}$ whenever $t \in [t_0 - 2\tau r_0^2, t_0]$ and $\mathrm{dist}_t(x,x_0) \leq \frac{1}{4}\,r_0$. In particular, $R(x,t) \leq K$ whenever $t \in [t_0 - \tau r_0^2, t_0]$ and $\mathrm{dist}_t(x,x_0) \leq \frac{1}{4}\,r_0$, which contradicts our assumption that the conclusion of the theorem fails.

Now suppose that $\tau' < 2\tau$. Put $t' = t_0 - \tau' r_0^2$. From estimates on the length and volume distortion under the Ricci flow, we know that there are numbers $\alpha = \alpha(w) > 0$ and $w' = w'(w) > 0$ so that the time-t' ball $B(x_0,\alpha r_0)$ has volume at least $w'(\alpha r_0)^3$. From Lemma 54.1, there is a subball $B(x',r') \subset B(x_0,\alpha r_0)$ with $r' \geq c\alpha r_0$ and $\mathrm{vol}(B(x',r')) \geq \frac{1}{2}\,\omega_3\,(r')^3$. From the preceding pointpicking argument, we have the estimate $R(x,t) \leq K(r')^{-2}$ whenever $t \in [t' - \tau(r')^2, t']$ and $\mathrm{dist}_t(x,x') \leq \frac{1}{4}r'$. From the Φ-almost nonnegative curvature, we have

a bound $|\operatorname{Rm}|(x,t) \leq$ const. $K\,(r')^{-2}$ + const. $\Phi(K(r')^{-2})$ at such a point (see (52.4)). If $\rho(w)$ is taken sufficiently small then we can ensure that r_0 is small enough, and hence r' is small enough, to make $K\,(r')^{-2} + \Phi(K(r')^{-2}) \leq 2\,K\,(r')^{-2}$. Then we can apply Theorem 53.1 to a time interval ending at time t', after a redefinition of its constants, to obtain a bound of the form $R(x,t') \leq K'(r')^{-2}$ whenever $\operatorname{dist}_t(x,x') \leq 10r_0$, where K' is related to the constant K of Theorem 53.1. (We also obtain a similar estimate at times slightly less than t'.) Thus at such a point, $\operatorname{Rm}(x,t') \geq -\Phi(K'(r')^{-2})$. If we choose $\rho(w)$ to be sufficiently small to force r' to be sufficiently small to force $-\Phi(K'(r')^{-2}) > -r_0^{-2}$ then we have $\operatorname{Rm} > -r_0^{-2}$ on $\overline{B_{t'}(x_0,r_0)} \subset B_{t'}(x',10r_0)$, which contradicts the assumed maximality of τ'.

We note that in the application of Theorem 53.1 at the end of the proof, we must take into account the extra hypothesis, in the notation of Theorem 53.1, that $r_0^2\,\Phi(r_0^{-2}) < \rho$ (see Remark 53.2). This will be satisfied if the r_0 in Theorem 53.1 is small enough, which is ensured by taking the ρ of Theorem 54.2 small enough. □

55. I.12.4. Small balls with strongly negative curvature are volume-collapsed

In this section we show that under certain hypotheses, if the infimal sectional curvature on an r-ball is exactly $-r^{-2}$ then the volume of the ball is small compared to r^3.

Corollary 55.1. *(cf. Corollary I.12.4) For any $w > 0$, one can find $\rho > 0$ with the following property. Suppose that $g(\cdot)$ is a Φ-almost nonnegatively curved Ricci flow solution on a closed three-manifold M, defined for $t \in [0,T)$ with $T \geq 1$. If $B(x_0,r_0)$ is a metric ball at time $t_0 \geq 1$ with $r_0 < \rho$ and if $\inf_{x\in B(x_0,r_0)} \operatorname{Rm}(x,t_0) = -r_0^{-2}$ then $\operatorname{vol}(B(x_0,r_0)) \leq wr_0^3$.*

Proof. Fix $w > 0$. The number ρ will be specified in the course of the proof. Suppose that the corollary is not true, i.e. there is a Ricci flow solution as in the statement of the corollary along with a metric ball $B(x_0,r_0)$ at a time $t_0 \geq 1$ so that $\inf_{x\in B(x_0,r_0)} \operatorname{Rm}(x,t_0) = -r_0^{-2}$ and $\operatorname{vol}(B(x_0,r_0)) > wr_0^n$. The idea is to use Theorem 54.2, along with the Φ-almost nonnegative curvature, to get a double-sided sectional curvature bound on a smaller ball at an earlier time. Then one goes forward in time using Theorem 53.1, along with the Φ-almost nonnegative curvature, to get a lower sectional curvature bound on the original ball, thereby obtaining a contradiction.

Looking at the hypotheses of Theorem 54.2, if we require $r_0 < (4\tau)^{-\frac{1}{2}}$ then $4\tau r_0^2 < 1 \leq t_0$. From Theorem 54.2, $R(x,t) \leq Kr_0^{-2}$ whenever $t \in [t_0 - \tau r_0^2, t_0]$ and $\operatorname{dist}_t(x,x_0) \leq \frac{1}{4}r_0$, provided that r_0 is small enough that $r_0^2\,\Phi(r_0^{-2})$ is less than the ρ of Theorem 54.2. If in addition r_0 is sufficiently small then it follows that $|\operatorname{Rm}(x,t)| \leq$ const. $\Phi(Kr_0^{-2}) \leq r_0^{-2}$.

From the Bishop-Gromov inequality and the bounds on length and volume distortion under Ricci flow, there is a small number c so that we are ensured that $|\operatorname{Rm}(x,t)| \leq (cr_0)^{-2}$ for all (x,t) satisfying $\operatorname{dist}_{t_0-(cr_0)^2}(x,x_0) < cr_0$ and $t \in [t_0 - (cr_0)^2, t_0]$, and in addition the volume of $B(x_0,cr_0)$ at time $t_0 - (cr_0)^2$ is at least $c(cr_0)^3$. Choosing the constant A of Theorem 53.1 appropriately in terms of c, we can apply Theorem 53.1 to the ball $B(x_0,cr_0)$ and the time interval $[t_0-(cr_0)^2,t_0]$ to conclude that at time t_0, $R(\cdot,t_0)\Big|_{B(x_0,r_0)} \leq$

$K(A)\,(cr_0)^{-2}$, where $K(A)$ is as in the statement of Theorem 53.1. From the Φ-almost nonnegative curvature condition,

$$\operatorname{Rm}\Big|_{B(x_0,r_0)} \geq -\Phi\left(K(A)\,(cr_0)^{-2}\right). \tag{55.2}$$

If r_0 is sufficiently small then we contradict the assumption that $\inf \operatorname{Rm}(x,t_0)\Big|_{B(x_0,r_0)} = -\frac{1}{r_0^2}$. ∎

56. I.13.1. Thick-thin decomposition for nonsingular flows

The main result of this section says that if $g(\cdot)$ is a Ricci flow solution on a closed oriented three-dimensional manifold M that exists for $t \in [0,\infty)$ then for large t, $(M, g(t))$ has a thick-thin decomposition. A fuller description is in Sections 87-92.

We assume that at time zero, the sectional curvatures are bounded below by -1. This can always be achieved by rescaling the initial metric. Then we have the Φ-almost nonnegative curvature result of (B.8).

If the metric $g(t)$ has nonnegative sectional curvature then it must be flat, as we are assuming that the Ricci flow exists for all time. Let us assume that $g(t)$ is not flat, so it has some negative sectional curvature. Given $x \in M$, consider the time-t ball $B_t(x,r)$. Clearly if r is sufficiently small then $\operatorname{Rm}\Big|_{B_t(x,r)} > -r^{-2}$, while if r is sufficiently large (maybe greater than the diameter of M) then it is not true that $\operatorname{Rm}\Big|_{B_t(x,r)} > -r^{-2}$. Let $\widehat{r}(x,t) > 0$ be the unique number such that $\inf \operatorname{Rm}\Big|_{B_t(x,\widehat{r})} = -\widehat{r}^{-2}$. Let $M_{thin}(w,t)$ be the set of points $x \in M$ for which

$$\operatorname{vol}(B_t(x,\widehat{r}(x,t))) < w\,\widehat{r}(x,t)^3. \tag{56.1}$$

Put $M_{thick}(w,t) = M - M_{thin}(w,t)$.

As the statement of (B.8) is invariant under parabolic rescaling (although we must take $t \geq t_0$ for (B.8) to apply), if $t \geq t_0$ and we are interested in the Ricci flow at time t then we can apply Theorem 53.1, Theorem 54.2 and Corollary 55.1 to the rescaled flow $g(t') = t^{-1} g(tt')$. From Corollary 55.1, for any $w > 0$ we can find $\widehat{\rho} = \widehat{\rho}(w) > 0$ so that if $\widehat{r}(x,t) < \widehat{\rho}\sqrt{t}$ then $x \in M_{thin}(w,t)$, provided that t is sufficiently large (depending on w). Equivalently, if t is sufficiently large (depending on w) and $x \in M_{thick}(w,t)$ then $\widehat{r}(x,t) \geq \widehat{\rho}\sqrt{t}$.

Theorem 56.2. *(cf. I.13.1) There are numbers $T = T(w) > 0$, $\overline{\rho} = \overline{\rho}(w) > 0$ and $K = K(w) < \infty$ so that if $t \geq T$ and $x \in M_{thick}(w,t)$ then $|\operatorname{Rm}| \leq Kt^{-1}$ on $B_t(x, \overline{\rho}\sqrt{t})$, and $\operatorname{vol}(B_t(x,\overline{\rho}\sqrt{t})) \geq \frac{1}{10}\,w\left(\overline{\rho}\sqrt{t}\right)^3$.*

Proof. The method of proof is the same as in Corollary 55.1. By assumption, $\operatorname{Rm}\Big|_{B_t(x,\widehat{r}(x,t))} \geq -\widehat{r}(x,t)^{-2}$ and $\operatorname{vol}(B_t(x,\widehat{r}(x,t))) \geq w\,\widehat{r}(x,t)^3$. As $\widehat{r}(x,t) \geq \widehat{\rho}\sqrt{t}$, for any $c \in (0,1)$ we have

$\operatorname{Rm}\Big|_{B_t(x,c\widehat{\rho}\sqrt{t})} \geq -(c\widehat{\rho})^{-2}t^{-1}$. By the Bishop-Gromov inequality,

$$\operatorname{vol}(B_t(x,c\widehat{\rho}\sqrt{t})) \geq \frac{\int_0^{\frac{c\widehat{\rho}\sqrt{t}}{\widehat{r}(x,t)}} \sinh^2(u)\,du}{\int_0^1 \sinh^2(u)\,du}\, w\,\widehat{r}(x,t)^3 \geq \frac{1}{3\int_0^1 \sinh^2(u)\,du}\, w\,(c\widehat{\rho})^3\, t^{\frac{3}{2}} \geq \frac{1}{10}\, w\,(c\widehat{\rho})^3\, t^{\frac{3}{2}}. \tag{56.3}$$

Considering Theorem 54.2 with its w replaced by $\frac{w}{10}$, if $c = c(w)$ is taken sufficiently small (to ensure $t \geq 4\tau(c\widehat{\rho}\sqrt{t})^2$) and t is larger than a certain w-dependent constant (to ensure $\frac{\Phi((c\widehat{\rho}\sqrt{t})^{-2})}{(c\widehat{\rho}\sqrt{t})^{-2}} < \rho$) then we can apply Theorem 54.2 with $r_0 = c\widehat{\rho}\sqrt{t}$ to obtain $R(x',t') \leq K'(w)c^{-2}\widehat{\rho}^{-2}t^{-1}$ whenever $t' \in [t - \tau c^2\widehat{\rho}^2 t, t]$ and $\operatorname{dist}_{t'}(x',x) \leq \frac{1}{4}\, c\widehat{\rho}\sqrt{t}$. From the Φ-almost nonnegative curvature (see (52.4)),

$$|\operatorname{Rm}|(x',t') \leq \text{const.}\ K'c^{-2}\widehat{\rho}^{-2}t^{-1} + \text{const.}\ \Phi(K'c^{-2}\widehat{\rho}^{-2}t^{-1}), \tag{56.4}$$

which is bounded above by 2 const. $K'c^{-2}\widehat{\rho}^{-2}t^{-1}$ if t is larger than a certain w-dependent constant. Then from length and volume distortion estimates for the Ricci flow, we obtain a lower volume bound $\operatorname{vol}(B_{t'}(x,c'\widehat{\rho}\sqrt{t})) \geq w'(c'\widehat{\rho}\sqrt{t})^3$ on a smaller ball of controlled radius, for some $c' = c'(w)$. Using Theorem 53.1, we finally obtain an upper bound $R \leq K''(w)(c\widehat{\rho}\sqrt{t})^{-2}$ on $B_t(x,c\widehat{\rho}\sqrt{t})$ and hence, by the Φ-almost nonnegative curvature, an upper bound of the form $|\operatorname{Rm}| \leq K(w)\,t^{-1}$ on $B_t(x,c\widehat{\rho}\sqrt{t})$, provided that $t \geq T$ for an appropriate $T = T(w)$. Taking $\overline{\rho} = c\widehat{\rho}$, the theorem follows. □

Remark 56.5. The use of Theorem 53.1 in the proof of Theorem 56.2 also gives an upper bound $|\operatorname{Rm}|(x,t) \leq K(A,w)\,t^{-1}$ on $B_t(x,A\overline{\rho}\sqrt{t})$ if $x \in M_{thick}(w,t)$, for any $A > 0$.

We now take w sufficiently small. Then for large t, $M_{thick}(w,t)$ has a boundary consisting of tori that are incompressible in M and the interior of $M_{thick}(w,t)$ admits a complete Riemannian metric with constant sectional curvature $-\frac{1}{4}$ and finite volume; see Sections 90 and 91. In addition, $M_{thin}(w,t)$ is a graph manifold; see Section 92.

57. Overview of *Ricci Flow with Surgery on Three-Manifolds* [52]

The paper [52] is concerned with the Ricci flow on compact oriented 3-manifolds. The main difference with respect to [51] is that singularity formation is allowed, so the paper deals with a "Ricci flow with surgery".

The main part of the paper is concerned with setting up the surgery procedure and showing that it is well-defined, in the sense that surgery times do not accumulate. In addition, the long-time behavior of a Ricci flow with surgery is analyzed.

The paper can be divided into three main parts. Sections II.1-II.3 contain preparatory material about ancient solutions, the so-called standard solution and the geometry at the first singular time. Sections II.4-II.5 set up the surgery procedure and prove that it is well-defined. Sections II.6-II.8 analyze the long-time behavior.

57.1. **II.1-II.3.** Section II.1 continues the analysis of three-dimensional κ-solutions from I.11. From I.11, any κ-solution contains an "asymptotic soliton", a gradient shrinking soliton that arises as a rescaled limit of the κ-solution as $t \to -\infty$. It is shown that any such gradient shrinking soliton must be a shrinking round cylinder $\mathbb{R} \times S^2$, its $\mathbb{Z}_2$-quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^2$ or a finite quotient of the round shrinking S^3. Using this, one obtains a finer description of the κ-solutions. In particular, any compact κ-solution must be isometric to a finite quotient of the round shrinking S^3, or diffeomorphic to S^3 or $\mathbb{R}P^3$. It is shown that there is a universal number $\kappa_0 > 0$ so that any κ-solution is a finite quotient of the round shrinking S^3 or is a κ_0-solution. This implies universal derivative bounds on the scalar curvature of a κ-solution.

Section II.2 defines and analyzes the Ricci flow of the so-called standard solution. This is a Ricci flow on $\mathbb{R}^3$ whose initial metric is a capped-off half cylinder. The surgery procedure will amount to gluing in a truncated copy of the time-zero slice of the standard solution. Hence one needs to understand the Ricci flow on the standard solution itself. It is shown that the Ricci flow of the standard solution exists on a maximal time interval $[0, 1)$, and the solution goes singular everywhere as $t \to 1$.

The geometry of the solution at the first singular time T (assuming that there is one) is considered in II.3. Put $\Omega = \{x \in M : \limsup_{t \to T^-} |\operatorname{Rm}(x, t)| < \infty\}$. Then Ω is an open subset of M, and $x \in M - \Omega$ if and only if $\lim_{t \to T^-} R(x, t) = \infty$. If $\Omega = \emptyset$ then for t slightly less than T, the manifold $(M, g(t))$ consists of nothing but high-scalar-curvature regions. Using Theorem I.12.1, one shows that M is diffeomorphic to $S^1 \times S^2$, $\mathbb{R}P^3 \# \mathbb{R}P^3$ or a finite isometric quotient of S^3.

If $\Omega \neq \emptyset$ then there is a well-defined limit metric $\overline{g}$ on Ω, with scalar curvature function $\overline{R}$. The set Ω could *a priori* have an infinite number of connected components, for example if an infinite number of distinct 2-spheres simultaneously shrink to points at time T. For small $\rho > 0$, put $\Omega_\rho = \{x \in \Omega : \overline{R}(x) \leq \rho^{-2}\}$, a compact subset of M. The connected components of Ω can be divided into those that intersect Ω_ρ and those that do not. If a connected component does not intersect Ω_ρ then it is a "capped ϵ-horn" (consisting of a hornlike end capped off by a ball or a copy of $\mathbb{R}P^3 - B^3$) or a "double ϵ-horn" (with two hornlike ends). If a connected component of Ω does intersect Ω_ρ then it has a finite number of ends, each being an ϵ-horn.

Topologically, the surgery procedure of II.4 will amount to taking each connected component of Ω that intersects Ω_ρ, truncating each of its ϵ-horns and gluing a 3-ball onto each truncated horn. The connected components of Ω that do not intersect Ω_ρ are thrown away. Call the new manifold M'. At a time t slightly less than T, the region $M - \Omega_\rho$ consists of high-scalar-curvature regions. Using the characterization of such regions in I.12.1, one shows that M can be reconstructed from M' by taking the connected sum of its connected components, along possibly with a finite number of $S^1 \times S^2$ and $\mathbb{R}P^3$ factors.

57.2. **II.4-II.5.** Section II.4 defines the surgery procedure. A Ricci flow with surgery consists of a sequence of smooth 3-dimensional Ricci flows on adjacent time intervals with the property that for any two adjacent intervals, there is a a compact 3-dimensional submanifold-with-boundary that is common to the final slice of the first time interval and the initial slice of the second time interval.

There are two *a priori* assumptions on a Ricci flow with surgery, the pinching assumption and the canonical neighborhood assumption. The pinching assumption is a form of Hamilton-Ivey pinching. The canonical neighborhood assumption says that every spacetime point (x, t) with $R(x, t) \geq r(t)^{-2}$ has a neighborhood which, after rescaling, is ϵ-close to one of the neighborhoods that occur in a κ-solution or in a time slice of the standard solution. Here ϵ is a small but universal constant and $r(\cdot)$ is a decreasing function, which is to be specified.

One wishes to define a Ricci flow with surgery starting from any compact oriented 3-manifold, say with a normalized initial metric. There are various parameters that will enter into the definition : the above canonical neighborhood scale $r(\cdot)$, a nonincreasing function $\delta(\cdot)$ that decays to zero, the truncation scale $\rho(t) = \delta(t)r(t)$ and the surgery scale h. In order to show that one can construct the Ricci flow with surgery, it turns out that one wants to perform the surgery only on necks with a radius that is very small compared to the canonical neighborhood scale; this is the role of the parameter $\delta(\cdot)$.

Suppose that the Ricci flow with surgery is defined at times less than T, with the *a priori* assumptions satisfied, and goes singular at time T. Define the open subset $\Omega \subset M$ as before and construct the compact subset $\Omega_\rho \subset M$ using $\rho = \rho(T)$. Any connected component N of Ω that intersects Ω_ρ has a finite number of ends, each of which is an ϵ-horn. This means that each point in the horn is in the center of an ϵ-neck, i.e. has a neighborhood that, after rescaling, is ϵ-close to a cylinder $\left[-\frac{1}{\epsilon}, \frac{1}{\epsilon}\right] \times S^2$. In II.4.3 it is shown that as one goes down the end of the horn, there is a self-improvement phenomenon; for any $\delta > 0$, one can find $h < \delta\rho$ so that if a point x in the horn has $\overline{R}(x) \geq h^{-2}$ then it is actually in the center of a δ-neck.

With $\delta = \delta(T)$, let h be the corresponding number. One then cuts off the ϵ-horn at a 2-sphere in the center of such a δ-neck and glues in a copy of a rescaled truncated standard solution. One does this for each ϵ-horn in N and each connected component N that intersects Ω_ρ, and throws away the connected components of Ω that do not intersect Ω_ρ. One lets the new manifold evolve under the Ricci flow. If one encounters another singularity then one again performs surgery. Based on an estimate on the volume change under a surgery, one concludes that a finite number of surgeries occur in any finite time interval. (However, one is not able to conclude from volume arguments that there is a finite number of surgeries altogether.)

The preceding discussion was predicated on the condition that the *a priori* assumptions hold for all times. For the Ricci flow before the first surgery time, the pinching condition follows from the Hamilton-Ivey result. One shows that surgery can be performed so that it does not make the pinching any worse. Then the pinching condition will hold up to the second surgery time, etc. The main issue is to show that one can choose the parameters $r(\cdot)$ and $\delta(\cdot)$ so that one knows *a priori* that the canonical neighborhood assumption, with parameter $r(\cdot)$, will hold for the Ricci flow with surgery. (For any singularity time T, one needs to know that the canonical neighborhood assumption holds for $t \in [0, T)$ in order to do the surgery at time T.)

As a preliminary step, in Lemma II.4.5 it is shown that after one glues in a standard solution, the result will still look similar to a standard solution, for as long of a time interval as one could expect, unless the entire region gets removed by some exterior surgery.

The result of II.5 is that the time-dependent parameters $r(\cdot)$ and $\delta(\cdot)$ can be chosen so as to ensure that the *a priori* assumptions hold. The normalization of the initial metric implies that there is a time interval $[0, C]$, for a universal constant C, on which the Ricci flow is smooth and has explicitly bounded curvature. On this time interval the canonical neighborhood assumption holds vacuously, if $r(\cdot)\big|_{[0,C]}$ is sufficiently small. To handle later times, the strategy is to divide $[\epsilon, \infty)$ into a countable sequence of finite time intervals and proceed by induction. In II.5 the intervals $\{[2^{j-1}\epsilon, 2^j\epsilon]\}_{j=1}^{\infty}$ are used, although the precise choice of intervals is immaterial.

We recall from I.12.1 that in the case of smooth flows, the proof of the canonical neighborhood assumption used the fact that one has κ-noncollapsing. It is not immediate that the method of proof of I.12.1 extends to a Ricci flow with surgery. (It is exactly for this reason that one takes $\delta(\cdot)$ to be a time-dependent function which can be forced to be very small.)

Hence one needs to prove κ-noncollapsing and the canonical neighborhoood assumption together. The main proposition of II.5 says that there are decreasing sequences r_j, κ_j and $\overline{\delta}_j$ so that if $\delta(\cdot)$ is a function with $\delta(\cdot)\big|_{[2^{j-1}\epsilon, 2^j\epsilon]} \leq \overline{\delta}_j$ for each $j > 0$ then any Ricci flow with surgery, defined with the parameters $r(\cdot)$ and $\delta(\cdot)$, is κ_j-noncollapsed on the time interval $[2^{j-1}\epsilon, 2^j\epsilon]$ at scales less than ϵ and satisfies the canonical neighborhood assumption there. Here we take $r(\cdot)\big|_{[2^{j-1}\epsilon, 2^j\epsilon)} = r_j$.

The proof of the proposition is by induction. Suppose that it is true for $1 \leq j \leq i$. In the induction step, besides defining the parameters r_{i+1}, κ_{i+1} and $\overline{\delta}_{i+1}$, one redefines $\overline{\delta}_i$. As one only redefines $\overline{\delta}$ in the previous interval, there is no circularity.

The first step of the proof, Lemma II.5.2, consists of showing that there is some $\kappa > 0$ so that for any r, one can find $\overline{\delta} = \overline{\delta}(r) > 0$ with the following property. Suppose that $g(\cdot)$ is a Ricci flow with surgery defined on $[0, T)$, with $T \in [2^i\epsilon, 2^{i+1}\epsilon]$, that satisfies the proposition on $[0, 2^i\epsilon]$. Suppose that it also satisfies the canonical neighborhood assumption with parameter r on $[2^i\epsilon, T)$, and is constructed using a function $\delta(\cdot)$ that satisfies $\delta(t) \leq \overline{\delta}$ on $[2^{i-1}\epsilon, T)$. Then it is κ-noncollapsed at all scales less than ϵ.

The proof of this lemma is along the lines of the κ-noncollapsing result of I.7, with some important modifications. One again considers the $\mathcal{L}$-length of curves $\gamma(\tau)$ starting from the point at which one wishes to prove the noncollapsing. One wants to find a spacetime point $(\overline{x}, \overline{t})$, with $\overline{t} \in [2^{i-1}\epsilon, 2^i\epsilon]$, at which one has an explicit upper bound on l. In I.7, the analogous statement came from a differential inequality for l. In order to use this differential equality in the present case, one needs to know that any curve $\gamma(\tau)$ that is competitive to be a minimizer for $L(\overline{x}, \overline{t})$ will avoid the surgery regions. Choosing $\overline{\delta}$ small enough, one can ensure that the surgeries in the time interval $[2^{i-1}\epsilon, T)$ are done on very long thin necks. Using Lemma II.4.5, one shows that a curve $\gamma(\tau)$ passing near such a surgery region obtains a large value of $\mathcal{L}$, thereby making it noncompetitive as a minimizer for $L(\overline{x}, \overline{t})$. (This is the underlying reason that the surgery parameter $\delta(\cdot)$ is chosen in a time-dependent way.) One

also chooses the point $(\overline{x}, \overline{t})$ so that there is a small parabolic neighborhood around it with a bound on its geometry. One can then run the argument of I.7 to prove κ-noncollapsing in the time interval $[2^i\epsilon, T)$.

The proof of the main proposition of II.5 is now by contradiction. Suppose that it is not true. Then for some sequences $r^\alpha \rightarrow 0$ and $\overline{\delta}^\alpha \rightarrow 0$, for each α there is a counterexample to the proposition with $r_{i+1} \le r^\alpha$ and $\delta_i, \delta_{i+1} \le \overline{\delta}^\alpha$. That is, there is some spacetime point in the interval $[2^i\epsilon, 2^{i+1}\epsilon]$ at which the canonical neighborhood assumption fails. Take a first such point (x^α, t^α). By Lemma II.5.2, one has κ-noncollapsing up to the time of this first counterexample. Using this noncollapsing, one can consider taking rescaled limits. If there are no surgeries in an appropriate-sized backward spacetime region around (x^α, t^α) then one can extract a convergent subsequence as $\alpha \rightarrow \infty$ and construct, as in the proof of Theorem I.12.1, a limit κ-solution, thereby giving a contradiction. If there are nearby interfering surgeries then one argues, using Lemma II.4.5, that the point (x^α, t^α) is in fact in a canonical neighborhood, again giving a contradiction.

Having constructed the Ricci flow with surgery, if the initial manifold is simply-connected then according to [24, 25, 53], there is a finite extinction time. One then concludes that the Poincaré Conjecture holds.

57.3. **II.6-II.8.** Sections II.6 and II.8 analyze the large-time behavior of a Ricci flow with surgery.

Section II.6 establishes back-and-forth curvature estimates. Proposition II.6.3 is an analog of Theorem I.12.2 and Proposition II.6.4 is an analog of Theorem I.12.3. The proofs are along the lines of the proofs of Theorems I.12.2 and I.12.3, but are complicated by the possible presence of surgeries.

The thick-thin decomposition for large-time slices is considered in Section II.7. Using monotonicity arguments of Hamilton, it is shown that as $t \rightarrow \infty$ the metric on the w-thick part $M^+(w, t)$ becomes closer and closer to having constant negative sectional curvature. Using a hyperbolic rigidity argument of Hamilton, it is stated that the hyperbolic pieces stabilize in the sense that there is a finite collection $\{(H_i, x_i)\}_{i=1}^k$ of pointed finite-volume 3-manifolds of constant sectional curvature $-\frac{1}{4}$ so that for large t, the metric $\widehat{g}(t) = \frac{1}{t}\, g(t)$ on the w-thick part $M^+(w, t)$ approaches the metric on the w-thick part of $\bigcup_{i=1}^k H_i$. It is stated that the cuspidal tori (if any) of the hyperbolic pieces are incompressible in M. To show this (following Hamilton), if there is a compressing 3-disk then one takes a minimal such 3-disk, say of area $A(t)$, and shows from a differential inequality for $A(\cdot)$ that for large t the function $A(t)$ is negative, which is a contradiction.

Theorem II.7.4, a statement in Riemannian geometry, characterizes the thin part $M^-(w, t)$, for small w and large t, as a graph manifold. The main hypothesis of the theorem is that for each point x, there is a radius $\rho = \rho(x)$ so that the ball $B(x, \rho)$ has volume at most $w\rho^3$ and sectional curvatures bounded below by $-\rho^{-2}$. In this sense the manifold is locally volume collapsed with respect to a lower sectional curvature bound.

Section II.8 contains an alternative proof of the incompressibility of cuspidal tori, using the functional $\lambda_1(g) = \lambda_1(-4\triangle + R)$. (At the beginning of Section 93, we give a simpler argument using the functional $R_{min}(g)\ \mathrm{vol}(M, g)^{\frac{2}{3}}$.) More generally, the functional $\lambda_1(g)$ is used to

define a topological invariant that determines the nature of the geometric decomposition. First, the manifold M admits a Riemannian metric g with $\lambda_1(g) > 0$ if and only if it admits a Riemannian metric with positive scalar curvature, which in turn is equivalent to saying that M is diffeomorphic to a connected sum of $S^1 \times S^2$'s and round quotients of S^3. If M does not admit a Riemannian metric with $\lambda_1 > 0$, let $\overline{\lambda}$ be the supremum of $\lambda_1(g) \cdot \mathrm{vol}(M, g)^{\frac{2}{3}}$ over all Riemannian metrics g on M. If $\overline{\lambda} = 0$ then M is a graph manifold. If $\overline{\lambda} < 0$ then the geometric decomposition of M contains a nonempty hyperbolic piece, with total volume $\left(-\frac{2}{3}\,\overline{\lambda}\right)^{\frac{3}{2}}$. The proofs of these statements use the monotonicity of $\lambda(g(t)) \cdot \mathrm{vol}(M, g(t))^{\frac{2}{3}}$, when it is nonpositive, under a smooth Ricci flow. The main work is to show that in the case of a Ricci flow with surgery, one can choose $\delta(\cdot)$ so that $\lambda(g(t)) \cdot \mathrm{vol}(M, g(t))^{\frac{2}{3}}$ is arbitrarily close to being nondecreasing in t.

58. II. Notation and terminology

$B(x, t, r)$ denotes the open metric ball of radius r, with respect to the metric at time t, centered at x.

$P(x, t, r, \Delta t)$ denotes a parabolic neighborhood, that is the set of all points (x', t') with $x' \in B(x, t, r)$ and $t' \in [t, t + \Delta t]$ or $t' \in [t + \Delta, t]$, depending on the sign of Δt.

Definition 58.1. We say that a Riemannian manifold (M_1, g_1) has distance $\leq \epsilon$ in the C^N-topology to another Riemannian manifold (M_2, g_2) if there is a diffeomorphism $\phi : M_2 \to M_1$ so that $\sum_{|I| \leq N} \frac{1}{|I|!} \parallel \nabla^I(\phi^* g_1 - g_2) \parallel_\infty \leq \epsilon$. An open set U in a Riemannian 3-manifold M is an *ϵ-neck* if modulo rescaling, it has distance less than ϵ, in the $C^{[1/\epsilon]+1}$-topology, to the product of the round 2-sphere of scalar curvature 1 (and therefore Gaussian curvature $\frac{1}{2}$) with an interval I of length greater than $2\epsilon^{-1}$. If a point $x \in M$ and a neighborhood U of x are specified then we will understand that "distance" refers to the pointed topology, where the basepoint in $S^2 \times I$ projects to the center of I.

We make a similar definition of ϵ-closeness in the spacetime case, where ∇^I now includes time derivatives. A subset of the form $U \times [a, b] \subset M \times [a, b]$, where $U \subset M$ is open, sitting in the spacetime of a Ricci flow is a *strong ϵ-neck* if after parabolic rescaling and time shifting, it has distance less than ϵ to the product Ricci flow defined on the time interval $[-1, 0]$ which, at its final time, is isometric to the product of a round 2-sphere of scalar curvature 1 with an interval of length greater than $2\epsilon^{-1}$. (Evidently, the time-0 slice of the product has 3-dimensional scalar curvature equal to 1.)

Our definition of an ϵ-neck differs in an insubstantial way from that on p. 1 of II. In the definition of [52], a ball $B(x, t, \epsilon^{-1} r)$ is called an ϵ-neck if, after rescaling the metric with a factor r^{-2}, it is ϵ-close, i.e. has distance less than ϵ, to *the corresponding subset of* the standard neck $S^2 \times I$... (italicized words added by us). (The issue is that a large metric ball in the cylinder $\mathbb{R} \times S^2$ does not have a smooth boundary.) Clearly after a slight change of the constants, an ϵ-neck in our sense is contained in an ϵ-neck in the sense of [52], and vice versa. An important fact is that the notion of (x, t) being contained in an ϵ-neck is an open condition with respect to the pointed $C^{[1/\epsilon]+1}$-topology on Ricci flow solutions.

With an ϵ-approximation $f : S^2 \times I \to U$ being understood, a cross-sectional sphere in U will mean the image of $S^2 \times \{\lambda\}$ under f, for some $\lambda \in (-\epsilon^{-1}, \epsilon^{-1})$. Any curve γ in U that intersects both $f(S^2 \times \{\epsilon^{-1}\})$ and $f(S^2 \times \{-\epsilon^{-1}\})$ must intersect each cross-sectional sphere. If γ is a minimizing geodesic and ϵ is small enough then γ will intersect each cross-sectional sphere exactly once.

There is a typo in the definition of a strong ϵ-neck in [52] : the parabolic neighborhood should be $P(x, t, \epsilon^{-1}r, -r^2)$, i.e. it should go backward in time rather than forward. We note that the time interval involved in the definition of strong ϵ-neck, i.e. 1 after rescaling, is different than the rescaled time interval ϵ^{-1} in Theorem 52.7.

In the next definition, $\mathbb{I}$ is an open interval and B^3 is an open ball.

Definition 58.2. A metric on $S^2 \times \mathbb{I}$ such that each point is contained in some ϵ-neck is called an *ϵ-tube*, or an *ϵ-horn*, or a *double ϵ-horn*, if the scalar curvature stays bounded on both ends, stays bounded on one end and tends to infinity on the other, or tends to infinity on both ends, respectively.

A metric on B^3 or $\mathbb{R}P^3 - \overline{B^3}$, such that each point outside some compact subset is contained in an ϵ-neck, is called an ϵ-cap or a capped ϵ-horn, if the scalar curvature stays bounded or tends to infinity on the end, respectively.

An example of an ϵ-tube is $S^2 \times (-\epsilon^{-1}, \epsilon^{-1})$ with the product metric. For a relevant example of an ϵ-horn, consider the metric

$$g = dr^2 + \frac{1}{8 \ln \frac{1}{r}} r^2 \, d\theta^2 \tag{58.3}$$

on $(0, R) \times S^2$, where $d\theta^2$ is the metric on S^2 with $R = 1$. From [6], the metric g models a rotationally symmetric neckpinch. Rescaling around r_0, we put $s = \sqrt{8 \ln \frac{1}{r_0}} \left(\frac{r}{r_0} - 1 \right)$ and find

$$\frac{8 \ln \frac{1}{r_0}}{r_0^2} g = ds^2 + \left(1 + \frac{1}{\sqrt{8 \ln \frac{1}{r_0}}} \left(2 + \frac{1}{\ln \frac{1}{r_0}} \right) s + O(s^2) \right) d\theta^2. \tag{58.4}$$

For small r_0, if we take $\epsilon \sim \left(\ln \frac{1}{r_0} \right)^{-\frac{1}{4}}$ then the region with $s \in (-\epsilon^{-1}, \epsilon^{-1})$ will be ϵ-biLipschitz close to the standard cylinder. Note that as $r_0 \to 0$, the constant ϵ improves; this is related to Lemma 71.1.

An ϵ-cap is the result of capping off an ϵ-tube by a 3-ball or $\mathbb{R}P^3 - \overline{B^3}$ with an arbitrary metric. A capped ϵ-horn is the result of capping off an ϵ-horn by a 3-ball or $\mathbb{R}P^3 - \overline{B^3}$ with an arbitrary metric.

Remark 58.5. Throughout the rest of these notes, ϵ denotes a small positive constant that is meant to be universal. The precise value of ϵ is unspecified. If the statement of a lemma or theorem invokes ϵ then the statement is meant to be true uniformly with respect to the other variables, provided ϵ is sufficiently small. When going through the proofs one is allowed to make ϵ small enough so that the arguments work, but one is only allowed to make a finite number of such reductions.

Lemma 58.6. *Let U be an ϵ-neck in an ϵ-tube (or horn) and let S be a cross-sectional sphere in U. Then S separates the two ends of the tube (or horn).*

Proof. Let W denote the tube (or horn). As any point $m \in W$ lies in some ϵ-neck, there is a unique lowest eigenvalue of the Ricci operator $\mathrm{Ric} \in \mathrm{End}(T_m W)$ at m. Let $\xi_m \subset T_m W$ be the corresponding eigenspace. As m varies, the ξ_m's form a smooth line field ξ on W, to which S is transverse.

Suppose that S does not separate the two ends of W. Then S represents a trivial element of $\mathrm{H}_2(S^2 \times I) \cong \pi_2(S^2 \times I)$ and there is an embedded 3-disk $D \subset W$ for which $\partial D = S$. This contradicts the fact that the line field ξ is transverse to S and extends over D. □

59. II.1. Three-dimensional κ-solutions

This section is concerned with properties of three-dimensional oriented κ-solutions. For brevity, in the rest of these notes we will generally omit the phrases "three-dimensional" and "oriented".

If $(M, g(\cdot))$ is a κ-solution then its topology is easy to describe. By definition, $(M, g(t))$ has nonnegative sectional curvature. If it does not have strictly positive curvature then the universal cover splits off a line (see Theorem A.7), from which it follows (using Corollary 40.1 and the κ-noncollapsing) that $(M, g(\cdot))$ is a standard shrinking cylinder $\mathbb{R} \times S^2$ or its $\mathbb{Z}_2$ quotient $\mathbb{R} \times_{\mathbb{Z}_2} S^2$. If $(M, g(t))$ has strictly positive curvature and M is compact then it is diffeomorphic to a spherical space form [35]. If $(M, g(t))$ has strictly positive curvature and M is noncompact then it is diffeomorphic to $\mathbb{R}^3$ [19]. The lemmas in this section give more precise geometric information. Recall that M_ϵ consists of the points in a κ-solution which are not the center of an ϵ-neck.

Lemma 59.1. *If $(M, g(t))$ is a time slice of a noncompact κ-solution and $M_\epsilon \neq \emptyset$ then there is a compact submanifold-with-boundary $X \subset M$ so that $M_\epsilon \subset X$, X is diffeomorphic to $\overline{B^3}$ or $\mathbb{R}P^3 - B^3$, and $M - \mathrm{int}(X)$ is diffeomorphic to $[0, \infty) \times S^2$.*

Proof. If $(M, g(t))$ does not have positive sectional curvature and $M_\epsilon \neq \emptyset$ then M must be isometric to $\mathbb{R} \times_{\mathbb{Z}_2} S^2$, in which case the lemma is easily seen to be true with X diffeomorphic to $\mathbb{R}P^3 - B^3$. Suppose that $(M, g(t))$ has positive sectional curvature. Choose $x \in M_\epsilon$. Let $\gamma : [0, \infty) \to M$ be a ray with $\gamma(0) = x$. As M_ϵ is compact, there is some $a > 0$ so that if $t > a$ then $\gamma(t) \notin M_\epsilon$. We can cover (a, ∞) by open intervals V_j so that $\gamma\big|_{V_j}$ is a geodesic segment in an ϵ-neck of rescaled length approximately $2\epsilon^{-1}$. Then we can find a cover of (a, ∞) by linearly ordered open intervals U_i, refining the previous cover, so that

1. The rescaled length of $\gamma\big|_{U_i}$ is approximately $\frac{1}{10}\,\epsilon^{-1}$.
2. Choosing some $x_i \in U_i \cap U_{i+1}$, the rescaled length (with rescaling at x_i) of $\gamma\big|_{U_i \cap U_{i+1}}$ is approximately $\frac{1}{40}\,\epsilon^{-1}$ and $\gamma\big|_{U_i \cap U_{i+1}}$ lies in an ϵ-neck W_i centered at x_i.

Let ϕ_i be projection on the first factor in the assumed diffeomorphism $W_i \cong (-\epsilon^{-1}, \epsilon^{-1}) \times S^2$. If ϵ is sufficiently small then the composition $\phi_i \circ \gamma\big|_{U_i} : U_i \to (-\epsilon^{-1}, \epsilon^{-1})$ is a diffeomorphism onto its image. Put $N_i = \phi_i^{-1}(\mathrm{Im}(\phi_i \circ \gamma\big|_{U_i})) \subset W_i$ and $p_i = (\phi_i \circ \gamma\big|_{U_i})^{-1} \circ \phi_i : N_i \to U_i \subset (a, \infty)$. Then p_i is ϵ-close to being a Riemannian submersion and on overlaps $N_i \cap N_{i+1}$, the maps p_i and p_{i+1} are C^K-close. Choosing an appropriate partition of unity $\{b_i\}$ subordinate to the U_i's, if ϵ is small then the function $f = \sum_i b_i p_i$ is a submersion from $\bigcup_i N_i$ to (a, ∞). The fiber is seen to be S^2. Given $t \in (a, \infty)$, put $X = M - f^{-1}(t, \infty)$. Then $M - \mathrm{int}(X) = f^{-1}([t, \infty))$ is diffeomorphic to $[0, \infty) \times S^2$.

Recall from Section 46 that we have an exhaustion of M by certain convex compact subsets. As M is one-ended, the subsets have connected boundary. As in Section 46, if the boundary of such a subset intersects an ϵ-neck then the intersection will be a nearly cross-sectional 2-sphere in the ϵ-neck. Hence with an appropriate choice of t, the set X will be isotopic to one of our convex subsets and so diffeomorphic to a closed 3-ball. □

Lemma 59.2. *If $(M, g(t))$ is a time slice of a κ-solution with $M_\epsilon = \emptyset$ then the Ricci flow is the evolving round cylinder $\mathbb{R} \times S^2$.*

Proof. By assumption, each point (x, t) lies in an ϵ-neck. If ϵ is sufficiently small then piecing the necks together, we conclude that M must be diffeomorphic to $S^1 \times S^2$ or $\mathbb{R} \times S^2$; see the proof of Lemma 59.1 for a similar argument. Then the universal cover $\widetilde{M}$ is $\mathbb{R} \times S^2$. As it has nonnegative sectional curvature and two ends, Toponogov's theorem implies that $(\widetilde{M}, \widetilde{g}(t))$ splits off an $\mathbb{R}$-factor. Using the strong maximum principle, the Ricci flow on $\widetilde{M}$ splits off an $\mathbb{R}$-factor; see Theorem A.7. Using Corollary 40.1, it follows that $(\widetilde{M}, \widetilde{g}(t))$ is the evolving round cylinder $\mathbb{R} \times S^2$. From the κ-noncollapsing, the quotient M cannot be $S^1 \times S^2$. □

A κ-solution has an asymptotic soliton (Section 39) that is either compact or noncompact. If the asymptotic soliton of a compact κ-solution $(M, g(\cdot))$ is also compact then it must be a shrinking quotient of the round S^3 [35], so the same is true of M.

Lemma 59.3. *If a κ-solution $(M, g(\cdot))$ is compact and has a noncompact asymptotic soliton then M is diffeomorphic to S^3 or $\mathbb{R}P^3$.*

Proof. We use Corollary 48.1 in Section 48. First, we claim that the time slices of the type-D κ-solutions of Corollary 48.1 have a universal upper bound on $\max_M R \cdot \mathrm{diam}(M)^2$. To see this, we can rescale at the point $x \in M_\epsilon$ by $R(x)$, after which the diameter is bounded above by $2\sqrt{\alpha}$. We then use Theorem 46.1 to get an upper bound on the rescaled scalar curvature, which proves the claim. Given an upper bound on $\max_M R \cdot \mathrm{diam}(M)^2$, the asymptotic soliton cannot be noncompact.

Thus we are in case C of Corollary 48.1. Take a sequence $t_i \to -\infty$ and choose points $x_i, y_i \in M_\epsilon(t_i)$ as in Corollary 48.1.C. Rescale by $R(x_i, t_i)$ and take a subsequence that converges to a pointed Ricci flow solution $(M_\infty, (x_\infty, t_\infty))$. The limit M_∞ cannot be compact, as otherwise we would have a uniform upper bound on $R \cdot \mathrm{diam}^2$ for $(M, g(t_i))$, which would contradict the existence of the noncompact asymptotic soliton. Thus M_∞ is a noncompact κ-solution. We can find compact sets $X_i \subset M$ containing $B(x_i, \alpha R(x_i, t_i)^{-\frac{1}{2}})$ so that $\{X_i\}$

converges to a set $X_\infty \subset M_\infty$ as in Lemma 59.1. Taking a further subsequence, we find similar compact sets $Y_i \subset M$ containing $B(y_i, \alpha R(y_i,t_i)^{-\frac{1}{2}})$ so that $\{Y_i\}$ converges to a set $Y_\infty \subset M_\infty$ as in Lemma 59.1. In particular, for large i, X_i and Y_i are each diffeomorphic to either $\overline{B^3}$ or $\mathbb{R}P^3 - B^3$. Considering a minimizing geodesic segment $\overline{x_i y_i}$ as in the statement of Corollary 48.1.C, we can use an argument as in the proof of Lemma 59.1 to construct a submersion from $M - (X_i \cup Y_i)$ to an interval, with fiber S^2. Hence M is diffeomorphic to the result of gluing X_i and Y_i along a 2-sphere. As M has finite fundamental group, no more than one of X_i and Y_i can be diffeomorphic to $\mathbb{R}P^3 - B^3$. Thus M is diffeomorphic to S^3 or $\mathbb{R}P^3$. □

From Lemma 50.1, every ancient solution which is a κ-solution *for some* κ is either a κ_0-solution or a metric quotient of the round S^3.

Lemma 59.4. *There is a universal constant η such that at each point of every ancient solution that is a κ-solution for some κ, we have estimates*

$$|\nabla R| < \eta R^{\frac{3}{2}}, \qquad |R_t| < \eta R^2. \tag{59.5}$$

Proof. This is obviously true for metric quotients of the round S^3. For κ_0-solutions it follows from the compactness result in Theorem 46.1, after rescaling the scalar curvature at the given point to be 1. □

It is sometimes useful to rewrite (59.5) as a pair of estimates on the spacetime derivatives of the quantity R^{-1} at points where $R \neq 0$:

$$|\nabla(R^{-\frac{1}{2}})| < \frac{\eta}{2}, \qquad |(R^{-1})_t| < \eta. \tag{59.6}$$

Lemma 59.7. *For every sufficiently small $\epsilon > 0$ one can find $C_1 = C_1(\epsilon)$ and $C_2 = C_2(\epsilon)$ such that for each point (x,t) in every κ-solution there is a radius $r \in [R(x,t)^{-1/2}, C_1 R(x,t)^{-1/2}]$ and a neighborhood B, $\overline{B(x,t,r)} \subset B \subset B(x,t,2r)$, which falls into one of the four categories:*
(a) B is a strong ϵ-neck (more precisely, B is the slice of a strong ϵ-neck at its maximal time, and an appropriate parabolic neighborhood of B satisfies the condition to be a strong ϵ-neck), or
(b) B is an ϵ-cap, or
(c) B is a closed manifold, diffeomorphic to S^3 or $\mathbb{R}P^3$, or
(d) B is a closed manifold of constant positive sectional curvature.

Furthermore:

- *The scalar curvature in B at time t is between $C_2^{-1} R(x,t)$ and $C_2 R(x,t)$.*
- *The volume of B in cases (a), (b) and (c) is greater than $C_2^{-1} R(x,t)^{-\frac{3}{2}}$.*
- *In case (b), there is an ϵ-neck $U \subset B$ with compact complement in B (i.e. the end of B is entirely contained in the ϵ-neck) such that the distance from x to U is at least $10000 R(x,t)^{-1/2}$.*
- *In case (c) the sectional curvature in B at time t is greater than $C_2^{-1} R(x,t)$.*

Remark 59.8. The statement of the lemma is slightly stronger than the corresponding statement in II.1.5, in that we have $r \geq R(x,t)^{-1/2}$ as opposed to $r > 0$.

Proof. We may assume that we are talking about a κ_0-solution, as if M is a metric quotient of a round sphere then it falls into category (d) for any $r > \pi(R(x_i, t_i)/6)^{-1/2}$ (since then $M = \overline{B(x_i, t_i, r)} = B(x_i, t_i, 2r)$).

Fix a small ϵ and suppose that the claim is not true. Then there is a sequence of κ_0-solutions M_i that together provide a counterexample. That is, there is a sequence $C_i \to \infty$ and a sequence of points $(x_i, t_i) \in M_i \times (-\infty, 0]$ so that for any $r \in [R(x_i, t_i)^{-1/2}, C_i R(x_i, t_i)^{-1/2}]$ one cannot find a B between $\overline{B(x_i, t_i, r)}$ and $B(x_i, t_i, 2r)$ falling into one of the four categories and satisfying the subsidiary conditions with parameter $C_2 = C_i$. Rescale the metric by $R(x_i, t_i)$ and take a convergent subsequence of $(M_i, (x_i, t_i))$ to obtain a limit κ_0-solution $(M_\infty, (x_\infty, t_\infty))$. Then for any $r > 1$, one cannot find a B_∞ between $\overline{B(x_\infty, t_\infty, r)}$ and $B(x_\infty, t_\infty, 2r)$ falling into one of the four categories and satisfying the subsidiary conditions for any parameter C_2.

If M_∞ is compact then for any r greater than the diameter of the time-t_∞ slice of M_∞, $\overline{B(x_\infty, t_\infty, r)} = M_\infty = B(x_\infty, t_\infty, 2r)$ falls into category (c) or (d). For the subsidiary conditions, M_∞ clearly has a lower volume bound, a positive lower scalar curvature bound and an upper scalar curvature bound. As a compact κ_0-solution has positive sectional curvature, M_∞ also has a lower sectional curvature bound. This is a contradiction.

If M_∞ is noncompact then Lemma 59.1 (or more precisely its proof) and Lemma 59.2 imply that for some $r > 1$, there will be a B between $\overline{B(x_\infty, t_\infty, r)}$ and $B(x_\infty, t_\infty, 2r)$ falling into category (a) or (b). In case (b), by choosing the parameter r sufficiently large, the existence of the ϵ-neck U with the desired properties follows from the proof of Lemma 59.1. For the other subsidiary conditions, B clearly has a lower volume bound, a positive lower scalar curvature bound and an upper scalar curvature bound. This contradiction completes the proof of the lemma. □

60. II.2. Standard solutions

The next few sections are concerned with the properties of special Ricci flow solutions on $M = \mathbb{R}^3$. We fix a smooth rotationally symmetric metric g_0 which is the result of gluing a hemispherical-type cap to a half-infinite cylinder of scalar curvature 1. Among other properties, g_0 is complete and has nonnegative curvature operator. We also assume that g_0 has scalar curvature bounded below by 1.

Remark 60.1. In Section 72 we will further specialize the initial metric g_0 of the standard solution, for technical convenience in doing surgeries.

Definition 60.2. A Ricci flow $(\mathbb{R}^3, g(\cdot))$ defined on a time interval $[0, a)$ is a *standard solution* if it has initial condition g_0, the curvature $|\operatorname{Rm}|$ is bounded on compact time intervals $[0, a'] \subset [0, a)$, and it cannot be extended to a Ricci flow with the same properties on a strictly longer time interval.

It will turn out that every standard solution is defined on the time interval $[0, 1)$.

To motivate the next few sections, let us mention that the surgery procedure will amount to gluing in a truncated copy of $(\mathbb{R}^3, g_0)$. The metric on this added region will then evolve as part of the Ricci flow that takes up after the surgery is performed. We will need to

understand the behavior of the Ricci flow after performing a surgery. Near the added region, this will be modeled by a standard solution. Hence one first needs to understand the Ricci flow of a standard solution.

The main results of II.2 concerning the Ricci flow on a standard solution (Sections 61-64) are used in II.4.5 (Lemma 74.1) to show that, roughly speaking, the part of the manifold added by surgery acquires a large scalar curvature soon after the surgery time. This is used crucially in II.5 (Sections 79 and 80) to adapt the noncollapsing argument of I.7 to Ricci flows with surgery.

Our order of presentation of the material in II.2 is somewhat different than that of [52]. In Sections 61-63, we cover Claims 2, 4 and 5 of II.2. These are what's needed in the sequel. The other results of II.2, Claims 1 and 3, are concerned with proving the uniqueness of the standard solution. Although it may seem intuitively obvious that there should be a unique and rotationally-symmetric standard solution, the argument is not routine since the manifold is noncompact.

In fact, the uniqueness is not really needed for the sequel. (For example, the method of proof of Lemma 74.1 produces *a* standard solution in a limiting argument and it is enough to know certain properties of this standard solution.) Because of this we will talk about *a* standard solution rather than *the* standard solution.

Consequently, we present the material so that we do not logically need the uniqueness of the standard solution. Having uniqueness does not shorten the subsequent arguments any. Of course, one can ask independently whether the standard solution is unique. In Section 65 we show that a standard solution is rotationally symmetric. In Section 66 we sketch the argument for uniqueness. Papers concerning the uniqueness of the standard solution are [21, 41].

We end this section by collecting some basic facts about standard solutions.

Lemma 60.3. *Let* $(\mathbb{R}^3, g(\cdot))$ *be a standard solution. Then*

(1) The curvature operator of g *is nonnegative.*

(2) All derivatives of curvature are bounded for small time, independent of the standard solution.

(3) The scalar curvature satisfies $\lim_{t\to a^-} \sup_{x\in\mathbb{R}^3} R(x,t) = \infty$.

(4) $(\mathbb{R}^3, g(\cdot))$ *is* κ*-noncollapsed at scales below* 1 *on any time interval contained in* $[0,2]$, *where* κ *depends only on the choice of the initial condition* g_0.

(5) $(\mathbb{R}^3, g(\cdot))$ *satisfies the conclusion of Theorem 52.7, in the sense that for any* $\Delta t > 0$, *there is an* $r_0 > 0$ *so that for any point* (x_0, t_0) *with* $t_0 \geq \Delta t$ *and* $Q = R(x_0, t_0) \geq r_0^{-2}$, *the solution in* $\{(x,t) : \mathrm{dist}^2_{t_0}(x, x_0) < (\epsilon Q)^{-1}, t_0 - (\epsilon Q)^{-1} \leq t \leq t_0\}$ *is, after scaling by the factor* Q, ϵ*-close to the corresponding subset of a* κ*-solution.*

Moreover, any Ricci flow which satisfies all of the conditions of Definition 60.2 except maximality of the time interval can be extended to a standard solution. In particular, using short-time existence [62, Theorem 1.1], *there is at least one standard solution.*

Proof. (1) follows from [63, Theorem 4.14].

(2) follows from Appendix D.

(3) In view of (1), this is equivalent to saying that $\lim_{t\to a^-}\sup_{x\in\mathbb{R}^3}|\mathrm{Rm}|(x,t)=\infty$. The argument for this last assertion is in [22, Chapter 6.7.2]. The proof in [22, Chapter 6.7.2] is for the compact case but using the derivative estimates of Appendix D, the same argument works in the present case.

(4) See Theorem 26.2.

(5) See Theorem 52.7.

The final assertion of the lemma follows from the method of proof of (3). □

61. Claim 2 of II.2. The blow-up time for a standard solution is ≤ 1

Lemma 61.1. *(cf. Claim 2 of II.2) Let $[0,T_S)$ be the maximal time interval such that the curvature of all standard solutions is uniformly bounded for every compact subinterval $[0,a]\subset[0,T_S)$. Then on the time interval $[0,T_S)$, the family of standard solution converges uniformly at (spatial) infinity to the standard Ricci flow on the round infinite cylinder $S^2\times\mathbb{R}$ of scalar curvature one. In particular, T_S is at most 1.*

Proof. Let $\{\mathcal{M}_i\}_{i=1}^\infty$ be a sequence of standard solutions, and let $\{x_i\}_{i=1}^\infty$ be a sequence tending to infinity in the time-zero slice M.

By (2) of Lemma 60.3, the gradient estimates in Appendix D, and Appendix E, every subsequence of $\{\mathcal{M}_i,(x_i,0)\}_{i=1}^\infty$ has a subsequence which converges in the pointed smooth topology on the time interval $[0,T_S)$. Therefore, it suffices to show that if $\{\mathcal{M}_i,(x_i,0)\}_{i=1}^\infty$ converges to some pointed Ricci flow $(\mathcal{M}_\infty,(x_\infty,0))$ then $\mathcal{M}_\infty$ is round cylindrical flow.

Since $g_i(0)=g_0$ for all i, the sequence of pointed time-zero slices $\{(M,x_i,g_i(0))\}_{i=1}^\infty$ converges in the pointed smooth topology to the round cylinder, i.e. $(M_\infty,g_\infty(0))$ is a round cylinder of scalar curvature 1. Each time slice $(M_\infty,g_\infty(t))$ is biLipschitz equivalent to $(M_\infty,g_\infty(0))$. In particular, it has two ends. As it also has nonnegative sectional curvature, Toponogov's theorem implies that $(M_\infty,g_\infty(t))$ splits off an $\mathbb{R}$-factor. Using the strong maximum principle, the Ricci flow $\mathcal{M}_\infty$ splits off an $\mathbb{R}$-factor; see Theorem A.7. Then using the uniqueness of the Ricci flow on the round S^2, it follows that $\mathcal{M}_\infty$ is a standard shrinking cylinder, which proves the lemma.

In particular, $T_S\leq 1$. □

62. Claim 4 of II.2. The blow-up time of a standard solution is 1

Lemma 62.1. *(cf. Claim 4 of II.2) Let T_S be as in Lemma 61.1. Then $T_S=1$. In particular, every standard solution survives until time 1.*

Proof. First, there is an $\alpha>0$ so that $T_S>\alpha$ [62, Theorem 1.1]. In what follows we will apply Theorem 52.7. The hypothesis of Theorem 52.7 says that the flow should exist on a time interval of duration at least one, but by rescaling we can apply Theorem 52.7 just as well with the alternative hypothesis that the flow exists on a time interval of duration at least α.

Suppose that $T_S < 1$. Then there is a sequence of standard solutions $\{\mathcal{M}_i\}_{i=1}^{\infty}$, times $t_i \to T_S$ and points $(x_i, t_i) \in \mathcal{M}_i$ so that $\lim_{i\to\infty} R(x_i, t_i) = \infty$.

We first argue that no subsequence of the points x_i can go to infinity (with respect to the time-zero slice). Suppose, after relabeling the subsequence, that $\{x_i\}_{i=1}^{\infty}$ goes to infinity. From Lemma 61.1, for any fixed $t' < T_S$ the pointed solutions $(M, (x_i, 0), g_i(\cdot))$, defined for $t \in [0, t']$, approach that of the shrinking cylinder on the same time interval. Lemma 60.3 and the characterization of high-curvature regions from Theorem 52.7 implies a uniform bound on high-curvature regions of the time derivative of R, of the form (59.5). Then taking t' sufficiently close to T_S, we get a contradiction. We conclude that outside of a compact region the curvature stays uniformly bounded as $t \to T_S$; compare with the proof of Lemma 52.11. (Alternatively, one could apply Theorem 30.1 to compact approximants, as is done in II.2.)

Thus we may assume that the sequence $\{x_i\}_{i=1}^{\infty}$ stays in a compact region of the time-zero slice. By Theorem 52.7, there is a sequence $\epsilon_i \to 0$ so that after rescaling the pointed solution $(\mathcal{M}, (x_i, t_i))$ by $R(x_i, t_i)$, the result is ϵ_i-close to the corresponding subset of an ancient solution. By Proposition 41.13, the ancient solutions have vanishing asymptotic volume ratio. Hence for every $\beta > 0$, there is some $L < \infty$ so that in the original unscaled solution, for large i we have $\mathrm{vol}\left(B(x_i, t_i, L\, R(x_i, t_i)^{-\frac{1}{2}})\right) \leq \beta \left(L\, R(x_i, t_i)^{-\frac{1}{2}}\right)^3$. Applying the Bishop-Gromov inequality to the time-t_i slices, we conclude that for any $D > 0$, $\lim_{i\to\infty} D^{-3}\,\mathrm{vol}(B(x_i, t_i, D)) = 0$. However, this contradicts the previously-shown fact that the solution extends smoothly to time $T_S < 1$ outside of a compact set.

Thus $T_S = 1$. $\square$

Lemma 62.2. *The infimal scalar curvature on the time-t slice tends to infinity as $t \to 1^-$ uniformly for all standard solutions.*

Proof. Suppose the lemma failed and let $\{(\mathcal{M}_i, (x_i, t_i))\}_{i=1}^{\infty}$ be a sequence of pointed standard solutions, with $\{R(x_i, t_i)\}_{i=1}^{\infty}$ uniformly bounded and $\lim_{i\to\infty} t_i = 1$.

Suppose first that after passing to a subsequence, the points x_i go to infinity in the time-zero slice. From Lemma 61.1, for any $t' \in [0, 1)$ we have $\lim_{i\to\infty} R^{-1}(x_i, t') = 1 - t'$. Combining this with the derivative estimate $\left|\frac{\partial R^{-1}}{\partial t}\right| \leq \eta$ at high curvature regions gives a contradiction; compare with the proof of Lemma 52.11. Thus the points x_i stay in a compact region. We can now use the bounded-curvature-at-bounded-distance argument in Step 2 of the proof of Theorem 52.7 to extract a convergent subsequence of $\{(\mathcal{M}_i, (x_i, 0))\}_{i=1}^{\infty}$ with a limit Ricci flow solution $(\mathcal{M}_\infty, (x_\infty, 0))$ that exists on the time interval $[0, 1]$. (In this case, the nonnegative curvature of the blowup region W comes from the fact that a standard solution has nonnegative curvature.) As in Step 3 of the proof of Theorem 52.7, $\mathcal{M}_\infty$ will have bounded curvature for $t \in [0, 1]$. Note that $\mathcal{M}_\infty$ is a standard solution. This contradicts Lemma 61.1. $\square$

63. Claim 5 of II.2. Canonical neighborhood property for standard solutions

Let p be the center of the hemispherical region in the time-zero slice.

Lemma 63.1. *(cf. Claim 5 of II.2) Given $\epsilon > 0$ sufficiently small, there are constants $\eta = \eta(\epsilon)$, $C_1 = C_1(\epsilon)$ and $C_2 = C_2(\epsilon)$ so that every standard solution $\mathcal{M}$ satisfies the conclusions of Lemmas 59.4 and 59.7, except that the ϵ-neck neighborhood need not be strong. (Here the constants do not depend on the standard solution.) More precisely, any point (x,t) is covered by one of the following cases :*

1.The time t lies in $(\frac{3}{4}, 1)$ and (x,t) has an ϵ-cap neighborhood or a strong ϵ-neck neighborhood as in Lemma 59.7.

2. $x \in B(p, 0, \epsilon^{-1})$, $t \in [0, \frac{3}{4}]$ and (x,t) has an ϵ-cap neighborhood as in Lemma 59.7.

3. $x \notin B(p, 0, \epsilon^{-1})$, $t \in [0, \frac{3}{4}]$ and there is an ϵ-neck $B(x, t, \epsilon^{-1} r)$ such that the solution in $P(x, t, \epsilon^{-1} r, -t)$ is, after scaling with the factor r^{-2}, ϵ-close to the appropriate piece of the evolving round infinite cylinder.

Moreover, we have an estimate $R_{min}(t) \geq \text{const.}\,(1-t)^{-1}$, where the constant does not depend on the standard solution.

Proof. We first show that the conclusion of Lemma 59.7 is satisfied.

In view of Lemma 62.2, there is a $\delta > 0$ so that if $t \in (1-\delta, 1)$ then we can apply Theorem 52.7 and Lemma 59.7 to a point (x,t) to see that the conclusions of Lemma 59.7 are satisfied in this case. If $t \in [0, 1-\delta]$ and x is sufficiently far from p (i.e. $\mathrm{dist}_0(x,p) \geq D$ for an appropriate D) then Lemma 61.1 implies that (x,t) has a strong ϵ-neck neighborhood or there is an ϵ-neck $B(x, t, \epsilon^{-1} r)$ such that the solution in $P(x, t, \epsilon^{-1} r, -t)$ is, after scaling with the factor r^{-2}, ϵ-close to the appropriate piece of the evolving round infinite cylinder.

(To elaborate a bit on the last possibility, the issue here is that there is no backward extension of the solution to $t < 0$. Because of this, if $t > 0$ is close to 0 then the backward neighborhood $P(x, t, \epsilon^{-1} r, -t)$ will not exist for rescaled time one, as required to have a strong ϵ-neck neighborhood. Since $\inf_{x \in M} R(x, 0) = 1$, we know from (B.2) that $R(x,t) \geq \frac{1}{1 - \frac{2}{3} t}$. Then if $t > \frac{3}{5}$, the time from the initial slice to (x,t), after rescaled by the scalar curvature, is bounded below by $t \, \frac{1}{1 - \frac{2}{3} t} > 1$. In particular, if $t \geq \frac{3}{4}$ then $r^2 t$ is at least one and we are ensured that the backward neighborhood $P(x, t, \epsilon^{-1} r, -t)$ does contain a strong ϵ-neck neighborhood.)

If $t \in [0, 1-\delta]$ and $\mathrm{dist}_0(x,p) < D$ then, provided that D and ϵ are chosen appropriately, we can say that (x,t) has an ϵ-cap neighborhood.

We now show that the conclusion of Lemma 59.4 is satisfied. If $t \in [1-\delta, 1)$ then the conclusion follows from Theorem 52.7 and Lemma 59.4. If $\delta' > 0$ is sufficiently small and $t \in [0, \delta']$ then the conclusion follows from Appendix D. If $t \in [\frac{1}{2}\,\delta', 1 - \frac{1}{2}\,\delta]$ then we have an upper scalar curvature bound from Lemma 62.1. From Hamilton-Ivey pinching (see Appendix B), this implies a double-sided sectional curvature bound. The conclusion of Lemma 59.4, when $t \in [\frac{1}{2}\,\delta', 1 - \frac{1}{2}\,\delta]$, now follows from the Shi estimates of Appendix D.

The last statement of the lemma follows from the estimate $\left| \frac{\partial R^{-1}}{\partial t} \right| \leq \text{const.}$, which holds for t near 1 (see Lemmas 59.4, 60.3(5) and 62.2) and then can be extended to all $t \in [0,1)$ (see Lemma 61). From Lemma 62.2, $\lim_{t \to 1} R^{-1}(x,t) = 0$ for every x. Thus

$R^{-1}(x,t) \leq$ const. $(1-t)$ for any (x,t). Equivalently,

$$R(x,t) \geq \text{const.}\,(1-t)^{-1}. \tag{63.2}$$

□

64. Compactness of the space of standard solutions

Lemma 64.1. *The family $\mathcal{ST}$ of pointed standard solutions $\{(\mathcal{M},(p,0))\}$ is compact with respect to pointed smooth convergence.*

Proof. This follows immediately from Appendix E and the fact that the constant T_S from Lemma 61.1 is equal to 1, by Lemma 62.1. □

65. Claim 1 of II.2. Rotational symmetry of standard solutions

Consider a standard solution $(M, g(\cdot))$. Since the time-zero metric g_0 is rotationally symmetric, it is clear by separation of variables that there is a rotationally symmetric solution for some time interval $[0,T)$. In this section we show that every standard solution is rotationally symmetric for each $t \in [0,1)$. Of course this would follow from the uniqueness of the standard solution; see [21, 41]. But the direct argument given here is the first step toward a uniqueness proof as in [41].

Lemma 65.1. *(cf. Claim 1 of II.2) Any Ricci flow solution in the space $\mathcal{ST}$ is rotationally symmetric for all $t \in [0,1)$.*

Proof. We first describe an evolution equation for vector fields which turns out to send Killing vector fields to Killing vector fields. Suppose that a vector field $u = \sum_m u^m \partial_m$ evolves by

$$u^m_t = u^m{}_{;}{}^k{}_k + R^m{}_i u^i. \tag{65.2}$$

Then

(65.3)
$$\begin{aligned}
\partial_t(u^m{}_{;i}) &= u^m_{t\,;i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= (u^m{}_{;}{}^k{}_k + R^m{}_k u^k)_{;\,i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= u^m{}_{;}{}^k{}_{ki} + R^m{}_{k;i} u^k + R^m{}_k u^k{}_{;i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= u^m{}_{;}{}^k{}_{ik} - R^m{}_{lki}\, u^l{}_{;}{}^k - R^k{}_{lki}\, u^m{}_{;}{}^l + R^m{}_{k;i} u^k + R^m{}_k u^k{}_{;i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= u^m{}_{;ki}{}^k - R^m{}_{lki}\, u^l{}_{;}{}^k - R_{li}\, u^m{}_{;}{}^l + R^m{}_{k;i} u^k + R^m{}_k u^k{}_{;i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= u^m{}_{;ik}{}^k - (R^m{}_{lki} u^l)_{;}{}^k - R^m{}_{lki}\, u^l{}_{;}{}^k - R_{ki}\, u^m{}_{;}{}^k + R^m{}_{k;i} u^k + R^m{}_k u^k{}_{;i} + (\partial_t \Gamma^m_{ki})\,u^k \\
&= u^m{}_{;ik}{}^k - R^m{}_{lki;}{}^k u^l - 2R^m{}_{lki}\, u^l{}_{;}{}^k - R_{ik}\, u^m{}_{;}{}^k + R^m{}_{k;i} u^k + R^m{}_k u^k{}_{;i} + (\partial_t \Gamma^m_{ki})\,u^k.
\end{aligned}$$

Contracting the second Bianchi identity gives

$$R_{mlki;}{}^k = R_{il;m} - R_{im;l}. \tag{65.4}$$

Also,

$$\partial_t \Gamma^m_{\ ki} = \partial_t (g^{ml}\Gamma_{lki}) = 2R^{ml}\Gamma_{lki} - g^{ml}(R_{lk,i} + R_{li,k} - R_{ik,l}) \tag{65.5}$$
$$= -R^m_{\ k;i} - R^m_{\ i;k} + R_{ik;}{}^m.$$

Substituting (65.4) and (65.5) in (65.3) gives

$$\partial_t (u^m_{\ ;i}) \ = \ u^m_{\ ;ik}{}^k - 2R^m_{\ lki}\, u^l_{\ ;}{}^k - R_{ik}\, u^m_{\ ;}{}^k + R^m_{\ k} u^k_{\ ;i}. \tag{65.6}$$

Then

$$\partial_t (u_{j;i}) \ = \ \partial_t (g_{jm} u^m_{\ ;i}) \ = \ -2R_{jm} u^m_{\ ;i} + g_{jm}\, \partial_t (u^m_{\ ;i}) \tag{65.7}$$
$$= \ u_{j;\,ik}{}^k - 2R_{jlki}\, u^l_{\ ;}{}^k - R_{ik}\, u_{j\,;}{}^k - R_{jk} u^k_{\ ;i}$$
$$= \ u_{j;\,ik}{}^k + 2R_{ikjl}\, u^l_{\ ;}{}^k - R_{ik}\, u_{j\,;}{}^k - R_{kj} u^k_{\ ;i}.$$

Equivalently, writing $v_{ij} = u_{j;i}$ gives

$$\partial_t v_{ij} \ = \ v_{ij;k}{}^k + 2R_i{}^k{}_j{}^l\, v_{kl} - R_i{}^k v_{kj} - R^k_{\ j} v_{ik}.$$

Then putting $L_{ij} = v_{ij} + v_{ji}$ gives

$$\partial_t L_{ij} \ = \ L_{ij;k}{}^k + 2R_i{}^k{}_j{}^l\, L_{kl} - R_i{}^k L_{kj} - R^k_{\ j} L_{ik}.$$

For any $\lambda \in \mathbb{R}$, we have

$$\partial_t (e^{2\lambda t}\, L_{ij} L^{ij}) \ = \ 2\lambda\, (e^{2\lambda t}\, L_{ij} L^{ij}) \ + \ (e^{2\lambda t} L_{ij} L^{ij})_{;k}{}^k - 2\, e^{2\lambda t}\, L_{ij;k}\, L^{ij;k} + Q(\mathrm{Rm}, e^{\lambda t} L), \tag{65.8}$$

where $Q(\mathrm{Rm}, L)$ is an algebraic expression that is linear in the curvature tensor Rm and quadratic in L. Putting $M_{ij} = e^{\lambda t} L_{ij}$ gives

$$\partial_t (M_{ij} M^{ij}) \ = \ 2\lambda\, M_{ij} M^{ij} \ + \ (M_{ij} M^{ij})_{;k}{}^k - 2\, M_{ij;k}\, M^{ij;k} + Q(\mathrm{Rm}, M). \tag{65.9}$$

Suppose that we have a Ricci flow solution $g(t)$, $t \in [0, T]$, with $g(0) \ = \ g_0$. Let $u(0)$ be a rotational Killing vector field for g_0. Let $u_\infty(0)$ be its restriction to (any) S^2, which we will think of as the 2-sphere at spatial infinity. Solve (65.2) for $t \in [0, T]$ with $u(t)$ bounded at spatial infinity for each t; due to the asymptotics coming from Lemma 61.1 (which is independent of the rotational symmetry question), there is no problem in doing so. Arguing as in the proof of Lemma 61.1, one can show that for any $t \in [0, T]$, at spatial infinity $u(t)$ converges to $u_\infty(0)$. Construct $M_{ij}(t)$ from $u(t)$. As $u(0)$ is a Killing vector field, $M_{ij}(0) = 0$. For any $t \in [0, T]$, at spatial infinity the tensor $M_{ij}(t)$ converges smoothly to zero. Suppose that λ is sufficiently negative, relative to the L^∞-norm of the sectional curvature on the time interval $[0, T]$. We can apply the maximum principle to (65.9) to conclude that $M_{ij}(t) = 0$ for all $t \in [0, T]$. Thus $u(t)$ is a Killing vector field for all $t \in [0, T]$.

To finish the argument, as Ben Chow pointed out, any Killing vector field u satisfies

$$u^m{}_{;}{}^k{}_k + R^m_{\ i} u^i \ = \ 0. \tag{65.10}$$

To see this, we use the Killing field equation to write

$$0 \ = \ u_{m;k}{}^k \ + \ u_{k;m}{}^k \ = \ u_{m;k}{}^k \ + \ u_{k;m}{}^k \ - \ u_{k;}{}^k{}_m \ = \ u_{m;}{}^k{}_k \ + \ u^k_{\ ;mk} \ - \ u^k_{\ ;km} \tag{65.11}$$
$$= \ u_{m;}{}^k{}_k \ - \ R^k_{\ imk}\, u^i \ = \ u_{m;}{}^k{}_k \ + \ R_{mi}\, u^i.$$

Then from (65.2), $u^m_t \ = \ 0$ and the Killing vector fields are not changing at all. This implies that $g(t)$ is rotationally symmetric for all $t \in [0, T]$. $\square$

66. Claim 3 of II.2. Uniqueness of the standard solution

In this section, which is not needed for the sequel, we outline an argument for the uniqueness of the standard solution. We do this for the convenience of the reader. Papers on the uniqueness issue are [21, 41]. Our argument is somewhat different than that of [52, Proof of Claim 3 of Section 2], which seems to have some unjustified statements.

In general, suppose that we have two Ricci flow solutions $\mathcal{M} \equiv (M, g(\cdot))$ and $\widehat{\mathcal{M}} \equiv (M, \widehat{g}(\cdot))$ with bounded curvature on each compact time interval and the same initial condition. We want to show that they coincide. As the set of times for which $g(t) = \widehat{g}(t)$ is closed, it suffices to show that it is relatively open. Thus it is enough to show that g and $\widehat{g}$ agree on $[0, T)$ for some small T.

We will carry out the Deturck trick in this noncompact setting, using a time-dependent background metric as in [5, Section 2]. The idea is to define a 1-parameter family of metrics $\{h(t)\}_{t\in[0,T)}$ by $h(t) = \phi^{-1}(t)^* g(t)$, where $\{\phi(t)\}_{t\in[0,T)}$ is a 1-parameter family of diffeomorphisms of M whose generator is the negative of the time-dependent vector field

$$W^i(t) \;=\; h^{jk} \left(\Gamma(h)^i{}_{jk} \;-\; \Gamma(\widehat{g})^i{}_{jk} \right), \tag{66.1}$$

with $\phi_0 = \mathrm{Id}$. More geometrically, as in [33, Section 6], we consider the solution of the harmonic heat flow equation $\frac{\partial F}{\partial t} = \triangle F$ for maps $F : M \to M$ between the manifolds $(M, g(t))$ and $(M, \widehat{g}(t))$, with $F(0) = \mathrm{Id}$.

We now specialize to the case when $(M, g(\cdot))$ and $(M, \widehat{g}(\cdot))$ come from standard solutions. The technical issue, which we do not address here, is to show that a solution to the harmonic heat flow will exist for some time interval $[0, T)$ with uniformly bounded derivatives; see [21]. One is allowed to use the asymptotics of Section 61 here and from Section 65, one can also assume that all of the metrics are rotationally invariant. In the rest of this section we assume the existence of such a solution F. By further reducing the time interval if necessary, we may assume that $F(t)$ is a diffeomorphism of M for each $t \in [0, T)$. Then $h(t) = F^{-1}(t)^* g(t)$. Clearly $h(0) = g(0) = \widehat{g}(0)$.

By Section 61, g and $\widehat{g}$ have the same spatial asymptotics, namely that of the shrinking cylinder. We claim that this is also true for h. That is, we claim that $(\mathcal{M}, h(\cdot))$ converges smoothly to the shrinking cylinder solution on $[0, T)$. It suffices to show that F converges smoothly to the identity on $[0, T)$. Suppose not. Let $\{x_i\}_{i=1}^\infty$ be a sequence of points in the time-zero slice so that no subsequence of the pointed spacetime maps $(F, (x_i, 0))$ converges to the identity. Using the derivative bounds, we can extract a subsequence that converges to some $\widetilde{F} : [0, T) \times \mathbb{R} \times S^2 \to \mathbb{R} \times S^2$ in the pointed smooth topology. However, $\widetilde{F}$ will satisfy the harmonic heat flow equation from the shrinking cylinder $\mathbb{R} \times S^2$ to itself, with $\widetilde{F}(0)$ being the identity, and will have bounded derivatives. The uniqueness of $\widetilde{F}$ follows by standard methods. Hence $\widetilde{F}(t)$ is the identity for all $t \in [0, T)$, which is a contradiction.

By construction, the family of metrics $\{h(t)\}_{t\in[0,T)}$ satisfies the equation

$$\frac{dh_{ij}}{dt} \;=\; -\,2\, R_{ij}(h) \;+\; \nabla(h)_i W_j \;+\; \nabla(h)_j W_i. \tag{66.2}$$

In local coordinates, the right-hand side of (66.2) is a polynomial in h_{ij}, h^{ij}, $h_{ij,k}$ and $h_{ij,kl}$. The leading term in (66.2) is

$$\frac{dh_{ij}}{dt} = h^{kl}\,\partial_k\partial_l h_{ij} + \dots . \tag{66.3}$$

A particular solution of (66.2) is $h(t) = \widehat{g}(t)$, since if we had $g = \widehat{g}$ then we would have $W = 0$ and $\phi_t = \mathrm{Id}$.

Put $w(t) = h(t) - \widehat{g}(t)$. We claim that w satisfies an equation of the form

$$\frac{dw}{dt} = -\nabla(\widehat{g})^*\nabla(\widehat{g})w + Pw + Qw, \tag{66.4}$$

where P is a first-order operator and Q is a zeroth-order operator. To obtain the leading derivative terms in (66.4), using (66.3) we write

$$\begin{aligned}\frac{dw_{ij}}{dt} &= h^{kl}\,\partial_k\partial_l h_{ij} - \widehat{g}^{kl}\,\partial_k\partial_l \widehat{g}_{ij} + \dots \\ &= \widehat{g}^{kl}\,\partial_k\partial_l w_{ij} + \left(h^{kl} - \widehat{g}^{kl}\right)\partial_k\partial_l h_{ij} + \dots \\ &= \widehat{g}^{kl}\,\partial_k\partial_l w_{ij} - \widehat{g}^{ka}\,w_{ab}\,h^{bl}\,\partial_k\partial_l h_{ij} + \dots \\ &= \widehat{g}^{kl}\,\partial_k\partial_l w_{ij} - \widehat{g}^{ka}\,h^{bl}\,\partial_k\partial_l h_{ij}\,w_{ab} + \dots\end{aligned} \tag{66.5}$$

A similar procedure can be carried out for the lower order terms, leading to (66.4). By construction the operators P and Q have smooth coefficients which, when expressed in terms of orthonormal frames, will be bounded on M. In fact, as h and $\widehat{g}$ have the same spatial asymptotics, it follows from [5, Proposition 4] that the operator on the right-hand side of (66.4) converges at spatial infinity to the Lichnerowicz Laplacian $\triangle_L(\widehat{g})$.

By assumption, $w(0) = 0$. We now claim that $w(t) = 0$ for all $t \in [0, T)$. Let $K \subset M$ be a codimension-zero compact submanifold-with-boundary. For any $\lambda \in \mathbb{R}$, we have

(66.6)

$$\begin{aligned}&e^{2\lambda t}\,\frac{d}{dt}\left(\frac{1}{2}\,e^{-2\lambda t}\int_K |w(t)|^2\,\mathrm{dvol}_{\widehat{g}(t)}\right) = \\ &\int_K \left[\left(-\lambda - \frac{R}{2}\right)|w|^2 + \langle w, \frac{dw}{dt}\rangle\right]\mathrm{dvol}_{\widehat{g}(t)} = \\ &\int_K \left[\left(-\lambda - \frac{R}{2}\right)|w|^2 - |\nabla(\widehat{g})w|^2 + w_{ab}P^{abcdi}\nabla(\widehat{g})_i w_{cd} + \langle w, Qw\rangle\right]\mathrm{dvol}_{\widehat{g}(t)} \pm \\ &\int_{\partial K}\langle w, \nabla_n w\rangle\,\mathrm{dvol}_{\partial\widehat{g}(t)} = \\ &\int_K \left[\left(-\lambda - \frac{R}{2}\right)|w|^2 - |\nabla(\widehat{g})_i w^{cd} - \frac{1}{2}P^{abcdi}w_{ab}|^2 + \frac{1}{4}|P^{abcdi}w_{ab}|^2 + \langle w, Qw\rangle\right]\mathrm{dvol}_{\widehat{g}(t)} \pm \\ &\int_{\partial K}\langle w, \nabla_n w\rangle\,\mathrm{dvol}_{\partial\widehat{g}(t)}\,.\end{aligned}$$

Choose

$$\lambda > \sup_{v \neq 0}\frac{\frac{1}{4}|P^{abcdi}v_{ab}|^2 + \langle v, Qv\rangle}{\langle v, v\rangle}. \tag{66.7}$$

On any subinterval $[0,T'] \subset [0,T)$, since w converges to zero at infinity and $(M, \widehat{g}(t))$ is standard at infinity, by choosing K appropriately we can make $\int_{\partial K} \langle w, \nabla_n w \rangle \, \mathrm{dvol}_{\partial \widehat{g}(t)}$ small. It follows that there is an exhaustion $\{K_i\}_{i=1}^{\infty}$ of M so that

$$e^{2\lambda t} \frac{d}{dt} \left(\frac{1}{2} \, e^{-2\lambda t} \int_{K_i} |w(t)|^2 \, \mathrm{dvol}_{\widehat{g}(t)} \right) \le \frac{1}{i} \tag{66.8}$$

for $t \in [0, T']$. Then

$$\int_{K_i} |w(t)|^2 \, \mathrm{dvol}_{\widehat{g}(t)} \le \frac{e^{2\lambda t} - 1}{\lambda i} \tag{66.9}$$

for all $t \in [0,T']$. Taking $i \to \infty$ gives $w(t) = 0$.

Thus $h = \widehat{g}$. From (66.1), $W = 0$ and so $h = g$. This shows that if $\mathcal{M}, \widehat{\mathcal{M}} \in \mathcal{ST}$ then $\mathcal{M} = \widehat{\mathcal{M}}$.

67. II.3. Structure at the first singularity time

This section is concerned with the structure of the Ricci flow solution at the first singular time, in the case when the solution does go singular.

Let M be a connected closed oriented 3-manifold. Let $g(\cdot)$ be a Ricci flow on M defined on a maximal time interval $[0,T)$ with $T < \infty$. One knows that $\lim_{t \to T^-} \max_{x \in M} |\operatorname{Rm}|(x,t) = \infty$.

From Theorem 26.2 and Theorem 52.7, given $\epsilon > 0$ there are numbers $r = r(\epsilon) > 0$ and $\kappa = \kappa(\epsilon) > 0$ so that for any point (x,t) with $Q = R(x,t) \ge r^{-2}$, the solution in $P(x,t,(\epsilon Q)^{-\frac{1}{2}}, (\epsilon Q)^{-1})$ is (after rescaling by the factor Q) ϵ-close to the corresponding subset of a κ-solution. By Lemma 59.4, the estimate (59.5) holds at (x,t), provided ϵ is sufficiently small. In addition, there is a neighborhood B of (x,t) as described in Lemma 59.7. In particular, B is a strong ϵ-neck, an ϵ-cap or a closed manifold with positive sectional curvature.

If M has positive sectional curvature at some time t then it is diffeomorphic to a finite quotient of the round S^3 and shrinks to a point at time T [35]. The topology of M satisfies the conclusion of the geometrization conjecture and M goes extinct in a finite time. Therefore for the remainder of this section we will assume that the sectional curvature does not become everywhere positive.

We now look at the behavior of the Ricci flow as one approaches the singular time T.

Definition 67.1. Define a subset Ω of M by

$$\Omega = \{x \in M : \sup_{t \in [0,T)} |\operatorname{Rm}|(x,t) < \infty\}. \tag{67.2}$$

Suppose that $x \in M - \Omega$, so there is a sequence of times $\{t_i\}$ in $[0,T)$ with $\lim_{i\to\infty} t_i = T$ and $\lim_{i \to \infty} |\operatorname{Rm}|(x,t_i) = \infty$. As $\min_M R(\cdot, t)$ in nondecreasing in t, the largest sectional curvature at (x, t_i) goes to infinity as $i \to \infty$. Then by the Φ-almost nonnegative sectional curvature result of Appendix B, $\lim_{i\to\infty} R(x,t_i) = \infty$. From the time-derivative estimate of (59.5), $\lim_{t \to T^-} R(x,t) = \infty$. Thus $x \in M - \Omega$ if and only if $\lim_{t\to T^-} R(x,t) = \infty$.

Lemma 67.3. *Ω is open in M.*

Proof. Given $x \in \Omega$, using the time-derivative estimate in (59.6) gives a bound of the form $|R(x,t)| \leq C$ for $t \in [0,T)$. Then using the spatial-derivative estimate in (59.6) gives a number $\widehat{r} > 0$ so that so that $|R(\cdot,t)| \leq 2C$ on $B(x,t,\widehat{r})$, for each $t \in [0,T)$. The Φ-almost nonnegative sectional curvature implies a bound of the form $|\operatorname{Rm}(\cdot,t)| \leq C'$ on $B(x,t,\widehat{r})$, for each $t \in [0,T)$. Then the length-distortion estimate of Section 27 implies that we can pick a neighborhood N of x so that $|\operatorname{Rm}| \leq C'$ on $N \times [0,T)$. Thus $N \subset \Omega$. □

Lemma 67.4. *Any connected component C of Ω is noncompact.*

Proof. Since M is connected, if C were compact then it would be all of M. This contradicts the assumption that there is a singularity at time T. □

We remark that *a priori*, the structure of $M - \Omega$ can be quite complicated. For example, it is not ruled out that an accumulating collection of 2-spheres in M can simultaneously shrink to points. That is, $M - \Omega$ could have a subset of the form $(\{0\} \cup \{\frac{1}{i}\}_{i=1}^{\infty}) \times S^2 \subset (-1,1) \times S^2$, the picture being that Ω contains a sequence of smaller and smaller adjacent double horns. One could even imagine a Cantor set's worth of 2-spheres simultaneously shrinking, although conceivably there may be additional arguments to rule out both of these cases.

Lemma 67.5. *If $\Omega = \emptyset$ then M is diffeomorphic to S^3, $\mathbb{R}P^3$, $S^1 \times S^2$ or $\mathbb{R}P^3 \# \mathbb{R}P^3$.*

Proof. The time-derivative estimate in (59.6) implies that for t slightly less than T, we have $R(x,t) \geq r^{-2}$ for all $x \in M$. Thus at that time, every $x \in M$ has a neighborhood that is in an ϵ-neck or an ϵ-cap, as described in Lemma 59.7. (Recall that we have already excluded the positively-curved case of the lemma.)

As in the proof of Lemma 59.1, by splicing together the projection maps associated with neck regions, one obtains an open subset $U \subset M$ and a 2-sphere fibration $U \to N$ where the fibers are nearly totally geodesic, and the complement of U is contained in a union of ϵ-caps. It follows that U is connected. If there are any ϵ-caps then there must be exactly two of them U_1, U_2, and they may be chosen to intersect U in connected open sets $V_i = U_i \cap U$ which are isotopic to product regions in both U and in the U_i's. The caps being diffeomorphic to $\overline{B^3}$ or $\mathbb{R}P^3 - B^3$, it follows that M is diffeomorphic to S^3, $\mathbb{R}P^3$ or $\mathbb{R}P^3 \# \mathbb{R}P^3$ if $U \neq M$; otherwise M is diffeomorphic to an S^2 bundle over a circle, and the orientability assumption implies that this bundle is diffeomorphic to $S^1 \times S^2$. □

In the rest of this section we assume that $\Omega \neq \emptyset$. From the local derivative estimates of Appendix D, there is a smooth Riemannian metric $\overline{g} = \lim_{t \to T^-} g(t)\Big|_{\Omega}$ on Ω. Let $\overline{R}$ denote its scalar curvature. Thus the scalar curvature function extends to a continuous function on the subset $(M \times [0,T)) \cup (\Omega \times \{T\}) \subset M \times [0,T]$.

Lemma 67.6. *$(\Omega, \overline{g})$ has finite volume.*

Proof. From the lower scalar curvature bound of (B.2) and the formula $\frac{d}{dt} \operatorname{vol}(M, g(t)) = - \int_M R \, \operatorname{dvol}_M$, we obtain an estimate of the form $\operatorname{vol}(M, g(t)) \leq \text{const.} + \text{const.}\, t^{\frac{3}{2}}$, for $t < T$. The lemma follows. □

Lemma 67.7. *There is a open neighborhood V of $(M - \Omega) \times \{T\}$ in $M \times [0, T]$ such that R^{-1} extends to a continuous function on V which vanishes on $(M - \Omega) \times \{T\}$.*

Proof. As observed above Lemma 67.3, $x \in M - \Omega$ if and only if $\lim_{t \rightarrow T^-} R^{-1}(x, t) = 0$. The lemma follows by applying (59.6) to suitable spacetime paths. □

Definition 67.8. For $\rho < \frac{r}{2}$, put $\Omega_\rho = \{x \in \Omega : \overline{R}(x) \leq \rho^{-2}\}$.

Lemma 67.9. *The function $\overline{R} : \Omega \rightarrow \mathbb{R}$ is proper; equivalently, if $\{x_i\} \subset \Omega$ is a sequence which leaves every compact subset of Ω, then $\lim_{i \rightarrow \infty} \overline{R}(x_i) = \infty$. In particular, Ω_ρ is a compact subset of M for every $\rho < r$.*

Proof. Suppose $\{x_i\} \subset \Omega$ is a sequence such that $\{\overline{R}(x_i)\}$ is bounded. After passing to a subsequence, we may assume that $\{x_i\}$ converges to some point $x_\infty \in M$. But R^{-1} is well-defined and continuous on V, and vanishes on $(M - \Omega) \times \{T\}$, so we must have $x_\infty \in \Omega$. □

We now consider the connected components of Ω according to whether they intersect Ω_ρ or not. First, let C be a connected component of Ω that does not intersect Ω_ρ. Given $x \in C$, there is a neighborhood B_x of x which is ϵ-close to a region as described in Lemma 59.7. From Lemma 67.4, the neighborhood B_x cannot be of type (c) or (d) in the terminology of Lemma 59.7.

We now introduce some terminology.

If a manifold Z is diffeomorphic to $\mathbb{R}^3$ or $\mathbb{R}P^3 - \overline{B^3}$ then any embedded 2-sphere $\Sigma \subset Z$ separates Z into two connected subsets, one of which has compact closure and the other contains the end of Z. We refer to the first component as the *compact side* and the other component as the *noncompact side.*

An open subset R of a Riemannian manifold is a *good cylinder* if:

- It is ϵ-close, modulo rescaling, to a segment of a round cylinder of scalar curvature 1.
- The diameter of R is approximately 100 times its cross-section.
- Every point in R, lies in an ϵ-neck in the ambient Riemannian manifold.

From Lemma 59.7, every ϵ-cap neighborhood B_x contains a good cylinder lying in the ϵ-neck at the end of B_x.

Lemma 67.10. *Suppose that for all $x \in C$, the neighborhood B_x can be taken to be a strong ϵ-neck as in case (a) of Lemma 59.7. Then C is a double ϵ-horn.*

Proof. Each point x has an ϵ-neck neighborhood. We can glue these ϵ-necks together to form a submersion from C to a 1-manifold, with fiber S^2; cf. the proof of Lemma 59.1. (We can do the gluing by successively adding on good cylinders, where the intersections of successive cylinders have diameter approximately 10 times the diameter of the cross-sections.) In view of Lemma 67.7, it follows in this case that C is a double ϵ-horn. □

Lemma 67.11. *Suppose that there is some $x \in C$ whose neighborhood B_x is an ϵ-cap as in case (a) of Lemma 59.7. Then C is a capped ϵ-horn.*

Proof. Put $p_1 = x$. Let R be a good cylinder in the ϵ-neck at the end of B_{p_1}. Now glue on successive good cylinders to R, as in the proof of the preceding lemma, going away from p_1.

Case 1 : Suppose this gluing process can be continued indefinitely. Then taking the union of B_{p_1} with the good cylinders, we obtain an open subset W of C which is diffeomorphic to $\mathbb{R}^3$ or $\mathbb{R}P^3 - \overline{B^3}$. We claim that W is a closed subset of Ω. If not then there is a sequence $\{x_k\}_{k=1}^{\infty} \subset W$ converging to some $x_\infty \in \Omega - W$. This implies that $\{\overline{R}(x_k)\}_{k=1}^{\infty}$ remains bounded. In view of the overlap condition between successive good cylinders, a subsequence of $\{x_k\}_{k=1}^{\infty}$ lies in an infinite number of mutually disjoint good cylinders, whose volumes have a positive lower bound (because of the upper scalar curvature bound at the points x_k). This contradicts Lemma 67.6.

Thus W is open and closed in Ω. Hence $W = C$ and we are done.

Case 2 : Now suppose that the gluing process cannot be continued beyond some good cylinder R_1. Then there must be a point $p_2 \in R_1$ such that B_{p_2} is an ϵ-cap. Also, note that the union W_1 of B_{p_1} with the good cylinders is diffeomorphic to $\mathbb{R}^3$ or $\mathbb{R}P^3 - \overline{B^3}$, and that R_1 has compact complement in W_1. Let $\Sigma \subset R_1$ be a cross-sectional 2-sphere passing through p_2.

We first claim that if V is the compact side of Σ in W_1, then V coincides with the compact side V' of Σ in B_{p_2}. To see this, note that V and V' are both connected open sets disjoint from Σ, with topological frontiers $\partial V = \partial V' = \Sigma$. Then $V - V' = V \cap (C - (V' \cup \Sigma))$ and we obtain two open decompositions

$$V = (V \cap V') \sqcup (V - V'), \quad V' = (V \cap V') \sqcup (V' - V). \tag{67.12}$$

If $V \cap V' = \emptyset$, then $\overline{V} \cup \overline{V'}$ is a union of two compact manifolds with the same boundary Σ, and disjoint interiors. Hence it is an open and closed subset of the connected component C, which contradicts Lemma 67.4. Thus $V \cap V'$ is nonempty. By (67.12) and the connectedness of V and V', we get $V \subset V'$ and $V' \subset V$, so $V = V'$ as claimed.

Next, we claim that if $R_2 \subset B_{p_2}$ is a good cylinder with compact complement in B_{p_2}, then R_2 is disjoint from W_1. To see this, note that R_2 is disjoint from Σ because $p_2 \in \Sigma$ and the diameter of Σ is close to $\pi(\overline{R}(p_2)/6)^{-1/2}$, whereas by Lemma 59.7 there is an ϵ-neck $U \subset B_{p_2}$ with compact complement in B_{p_2}, at distance at least $9000\overline{R}(p_2)^{-1/2}$ from p_2. Thus R_2 must lie in the noncompact side of Σ in B_{p_2}, and hence is disjoint from V. As the good cylinder $R_1 \ni p_2$ lies within $B(p_2, 1000\overline{R}(p_2)^{-1/2}) \subset \Omega$, it follows that R_2 is also disjoint from R_1, so R_2 is disjoint from $W_1 = V \cup R_1 \subset B_{p_2}$.

We continue adding good cylinders to R_2 as long as we can. If we come to another cap point p_3 then we jump to its cap B_{p_3} and continue the process. When so doing, we encounter successive cap points $p_1, p_2, \ldots$ with associated caps $B_{p_1} \subset B_{p_2} \subset \ldots$ and disjoint good cylinders $R_1, R_2, \ldots$. Since the ratio $\frac{\sup_{B_{p_k}} \overline{R}}{\inf_{B_{p_k}} \overline{R}}$ has an *a priori* bound by Lemma 59.7, in view of the disjoint good cylinders in B_{p_k} we get $\mathrm{vol}(B_{p_k}) \geq \mathrm{const.}\, k\, \overline{R}(p_1)^{-3/2}$. Then

Lemma 67.6 gives an upper bound on k. Hence we encounter a finite number of cap points. Arguing as in Case 1, we conclude that C is a capped ϵ-horn. $\square$

We note that there could be an infinite number of connected components of Ω that do not intersect Ω_ρ.

Now suppose that C is a connected component of Ω that intersects Ω_ρ. As C is noncompact, there must be some point $x \in C$ that is not in Ω_ρ. Again, any such x has a neighborhood B as in Lemma 59.7. If one of the boundary components of B intersects $\Omega_{2\rho}$ then we terminate the process in that direction. For the directions of the boundary components of B that do not intersect $\Omega_{2\rho}$, we perform the above algorithm of looking for an adjacent ϵ-neck, etc. The only difference from before is that in at least one direction any such sequence of overlapping ϵ-necks will be finite, as it must eventually intersect $\Omega_{2\rho}$. (In the other direction it may terminate in $\Omega_{2\rho}$, in an ϵ-cap, or not terminate at all.) Once a cross-sectional 2-sphere intersects $\Omega_{2\rho}$, if ϵ is small then the entire 2-sphere lies in Ω_ρ. Thus any connected component of $C - (C \cap \Omega_\rho)$ is contained in an ϵ-tube with both boundary components in Ω_ρ, an ϵ-cap with boundary in Ω_ρ or an ϵ-horn with boundary in Ω_ρ. We note that Ω_ρ need not have a nice boundary.

There is an *a priori* ρ-dependent lower bound for the volume of any such connected component of $C-(C\cap\Omega_\rho)$, in view of the fact that it contains ϵ-necks that adjoin Ω_ρ. From Lemma 67.6, there is a finite number of connected components of Ω that intersect Ω_ρ. Any such connected component has a finite number of ends, each being an ϵ-horn. Note that the ϵ-horns can be made disjoint, each with a quantitative lower volume bound.

The surgery procedure, which will be described in detail in Section 73, is performed as follows. First, one throws away all connected components of Ω that do not intersect Ω_ρ. For each connected component Ω_j of Ω that intersects Ω_ρ and for each ϵ-horn of Ω_j, take a cross-sectional sphere that lies far in the ϵ-horn. Let X be what's left after cutting the ϵ-horns at these 2-spheres and removing the tips. The (possibly-disconnected) postsurgery manifold M' is the result of capping off ∂X by 3-balls.

We now discuss how to reconstruct the original manifold M from M'.

Lemma 67.13. *M is the result of taking connected sums of components of M' and possibly taking additional connected sums with a finite number of $S^1 \times S^2$'s and $\mathbb{R}P^3$'s.*

Proof. At a time shortly before T, each point of $M - X$ has a neighborhood as in Lemma 59.7. The components of $M - X$ are ϵ-tubes and ϵ-caps. Writing $M' = X \cup \bigcup B^3$ and $M = X \cup (M - X)$, one builds M from M' as follows. If the boundary of an ϵ-tube of $M - X$ lies in two disjoint components of X then it gives rise to a connected sum of two components of M'. If the boundary of an ϵ-tube lies in a single connected component of X then it gives rise to the connected sum of the corresponding component of M' with a new copy of $S^1 \times S^2$. If an ϵ-cap in $M - X$ is a 3-ball it does not have any effect on M'. If an ϵ-cap is $\mathbb{R}P^3 - B^3$ then it gives rise to the connected sum of the corresponding component of M' with a new copy of $\mathbb{R}P^3$. The lemma follows. $\square$

Remark 67.14. We do not assume that the diameter of $(M, g(t))$ stays bounded as $t \to T$; it is an open question whether this is the case.

68. Ricci flow with surgery: the general setting

In this section we introduce some notation and terminology in order to treat Ricci flows with surgery.

The principal purpose of sections II.4 and II.5 is to show that one can prescribe the surgery procedure in such a way that Ricci flow with surgery is well-defined for all time. This involves showing that

- One can give a sufficiently precise description of the formation of singularities so that one can envisage defining a geometric surgery. In the case of the formation of the first singularity, such a description was given in Section 67.
- The sequence of surgery times cannot accumulate.

The argument in Section 67 strongly uses both the κ-noncollapsing result of Theorem 26.2 and the characterization of the geometry in a spacetime region around a point (x_0, t_0) with large scalar curvature, as given in Theorem 52.7. The proofs of both of these results use the smoothness of the solution at times before t_0. If surgeries occur before t_0 then one must have strong control on the scales at which the surgeries occur, in order to extend the arguments of Theorems 26.2 and 52.7. This forces one to consider time-dependent scales.

Section II.4 introduces Ricci flow with surgery, in varying degrees of generality. Our treatment of this material follows Perelman's. We have added some terminology to help formalize the surgery process. There is some arbitrariness in this formalization, but the version given below seems adequate.

For later use, we now summarize the relevant notation that we introduce. More precise definitions will be given below. We will avoid using new notation as much as possible.

- $\mathcal{M}$ is a Ricci flow with surgery.
- $\mathcal{M}_t$ is the time-t slice of $\mathcal{M}$.
- $\mathcal{M}_{\mathrm{reg}}$ is the set of regular points of $\mathcal{M}$.
- If T is a singular time then M_T^- is the limit of time slices $\mathcal{M}_t$ as $t \to T^-$ (called Ω in II.4.1) and M_T^+ is the outgoing time slice (for example, the result of performing surgery on Ω). If T is a nonsingular time then $\mathcal{M}_T^- = \mathcal{M}_T^+ = \mathcal{M}_T$.

The basic notion of a Ricci flow with surgery is simply a sequence of Ricci flows which "fit together" in the sense that the final (possibly singular) time slice of each flow is isometric, modulo surgery, to the initial time slice of the next one.

Definition 68.1. A *Ricci flow with surgery* is given by

- A collection of Ricci flows $\{(M_k \times [t_k^-, t_k^+), g_k(\cdot))\}_{1 \le k \le N}$, where $N \le \infty$, M_k is a compact (possibly empty) manifold, $t_k^+ = t_{k+1}^-$ for all $1 \le k < N$, and the flow g_k goes singular at t_k^+ for each $k < N$. We allow t_N^+ to be ∞.
- A collection of limits $\{(\Omega_k, \bar{g}_k)\}_{1 \le k \le N}$, in the sense of Section 67, at the respective final times t_k^+ that are singular if $k < N$. (Recall that Ω_k is an open subset of M_k.)
- A collection of isometric embeddings $\{\psi_k : X_k^+ \to X_{k+1}^-\}_{1 \le k < N}$ where $X_k^+ \subset \Omega_k$ and $X_{k+1}^- \subset M_{k+1}$, $1 \le k < N$, are compact 3-dimensional submanifolds with boundary. The

$X_k^{\pm}$'s are the subsets which survive the transition from one flow to the next, and the ψ_k's give the identifications between them.

We will say that t is a *singular time* if $t = t_k^+ = t_{k+1}^-$ for some $1 \le k < N$, or $t = t_N^+$ and the metric goes singular at time t_N^+.

A Ricci flow with surgery does not necessarily have to have any real surgeries, i.e. it could be a smooth nonsingular flow. Our definition allows Ricci flows with surgery that are more general than those appearing in the argument for geometrization, where the transitions/surgeries have a very special form. Before turning to these more special flows in Section 73, we first discuss some basic features of Ricci flow with surgery.

It will be convenient to associate a (non-manifold) spacetime $\mathcal{M}$ to the Ricci flow with surgery. This is constructed by taking the disjoint union of the smooth manifolds-with-boundary

$$\big(M_k \times [t_k^-, t_k^+)\big) \cup \big(\Omega_k \times \{t_k^+\}\big) \subset M_k \times [t_k^-, t_k^+] \tag{68.2}$$

for $1 \le k \le N$ and making identifications using the ψ_k's as gluing maps. We denote the quotient space by $\mathcal{M}$ and the quotient map by π. We will sometimes also use $\mathcal{M}$ to refer to the whole Ricci flow with surgery structure, rather than just the associated spacetime. The *time-t slice $\mathcal{M}_t$ of $\mathcal{M}$* is the image of the time-t slices of the constituent Ricci flows under the quotient map.

If $t = t_k^+$ is a singular time then we put $\mathcal{M}_t^- = \pi(\Omega_k \times \{t_k^+\})$; if in addition $t \ne t_N^+$ then we put $\mathcal{M}_t^+ = \pi(M_{k+1} \times \{t_{k+1}^-\})$. If t is not a singular time then we put $\mathcal{M}_t^+ = \mathcal{M}_t^- = \mathcal{M}_t$. We refer to $\mathcal{M}_t^+$ and $\mathcal{M}_t^-$ as the forward and backward time slices, respectively.

Let us summarize the structure of $\mathcal{M}$ near a singular time $t = t_k^+ = t_{k+1}^-$. The backward time slice $\mathcal{M}_t^-$ is a copy of Ω_k. The forward time slice $\mathcal{M}_t^+$ is a copy of M_{k+1}. The time slice $\mathcal{M}_t$ is the result of gluing Ω_k and M_{k+1} using ψ_k. Thus it is the disjoint union of $\Omega_k - X_k^+$, $M_{k+1} - X_{k+1}^-$ and $X_k^+ \cong X_{k+1}^-$. If $s > 0$ is small then in going from M_{t-s} to M_{t+s}, the topological change is that we remove $M_k - X_k^+$ from M_k and add $M_{k+1} - X_{k+1}^-$.

We let $\mathcal{M}_{(t,t')} = \cup_{\bar{t} \in (t,t')} \mathcal{M}_{\bar{t}}$ denote the *time slab* between t and t', i.e. the union of the time slices between t and t'. The closed time slab $\mathcal{M}_{[t,t']}$ is defined to be the closure of $\mathcal{M}_{(t,t')}$ in $\mathcal{M}$, so $\mathcal{M}_{[t,t']} = \mathcal{M}_t^+ \cup \mathcal{M}_{(t,t')} \cup \mathcal{M}_{t'}^-$. We (ab)use the notation (x,t) to denote a point $x \in \mathcal{M}$ lying in the time t slice $\mathcal{M}_t$, even though $\mathcal{M}$ may no longer be a product.

The spacetime $\mathcal{M}$ has three types of points:
1. The 4-manifold points, which include all points at nonsingular times in (t_1^-, t_N^+) and all points in $\pi(\mathrm{Interior}(X_k^+) \times \{t_k^+\})$ (or $\pi(\mathrm{Interior}(X_k^-) \times \{t_k^-\})$) for $1 \le k < N$,
2. The boundary points of $\mathcal{M}$, which are the images in $\mathcal{M}$ of $M_1 \times \{t_1^-\}$, $\Omega_N \times \{t_N^+\}$, $(\Omega_k - X_k^+) \times \{t_k^+\}$ for $1 \le k < N$, and $(M_k - X_k^-) \times \{t_k^-\}$ for $1 < k \le N$, and
3. The "splitting" points, which are the images in $\mathcal{M}$ of $\partial X_k^+ \times \{t_k^+\}$ for $1 \le k < N$.

Here the classification of points is according to the smooth structure, not the topology. In fact, $\mathcal{M}$ is a topological manifold-with-boundary. We say that (x,t) is *regular* if it is either a 4-manifold point, or it lies in the initial time slice $\mathcal{M}_{t_1^-}$ or final time slice $\mathcal{M}_{t_N^+}$. Let $\mathcal{M}_{\mathrm{reg}}$ denote the set of regular points. It has a natural smooth structure since the gluing maps ψ_k, being isometries between smooth Riemannian manifolds, are smooth maps.

Note that the Ricci flows on the M_k's define a Riemannian metric g on the "horizontal" subbundle of the tangent bundle of $\mathcal{M}_{\mathrm{reg}}$. It follows from the definition of the Ricci flow that g is actually smooth on $\mathcal{M}_{\mathrm{reg}}$.

We metrize each time slice $\mathcal{M}_t$, and the forward and backward time slices $\mathcal{M}_t^{\pm}$, by infimizing the path length of piecewise smooth paths. We allow our distance functions to be infinite, since the infimum will be infinite when points lie in different components. If $(x,t) \in \mathcal{M}_t$ and $r > 0$ then we let $B(x,t,r)$ denote the corresponding metric ball. Similarly, $B^{\pm}(x,t,r)$ denotes the ball in $\mathcal{M}_t^{\pm}$ centered at $(x,t) \in \mathcal{M}_t^{\pm}$. A ball $B(x,t,r) \subset \mathcal{M}_t$ is *proper* if the distance function $d_{(x,t)} : B(x,t,r) \to [0,r)$ is a proper function; a proper ball "avoids singularities", except possibly at its frontier. Proper balls $B^{\pm}(x,t,r) \subset \mathcal{M}_t^{\pm}$ are defined likewise.

An *admissible curve* in $\mathcal{M}$ is a path $\gamma : [c,d] \to \mathcal{M}$, with $\gamma(t) \in \mathcal{M}_t$ for all $t \in [c,d]$, such that for each k, the part of γ landing in $\mathcal{M}_{[t_k^- t_k^+]}$ lifts to a smooth map into $M_k \times [t_k^-, t_k^+) \cup \Omega_k \times \{t_k^+\}$. We will use $\dot{\gamma}$ to denote the "horizontal" part of the velocity of an admissible curve γ. If $t < t_0$, a point $(x,t) \in \mathcal{M}$ is *accessible from* $(x_0,t_0) \in \mathcal{M}$ if there is an admissible curve running from (x,t) to (x_0,t_0). An admissible curve $\gamma : [c,d] \to \mathcal{M}$ is *static* if its lifts to the product spaces have constant first component. That is, the points in the image of a static curve are "the same", modulo the passage of time and identifications taking place at surgery times. A *barely admissible curve* is an admissible curve $\gamma : [c,d] \to \mathcal{M}$ such that the image is not contained in $\mathcal{M}_c^+ \cup \mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}_d^-$. If $\gamma : [c,d] \to \mathcal{M}$ is barely admissible then there is a surgery time $t = t_k^+ = t_{k+1}^- \in (c,d)$ such that $\gamma(t)$ lies in

$$\pi(\partial X_k^+ \times \{t_k^+\}) = \pi(\partial X_{k+1}^- \times \{t_{k+1}^-\}). \tag{68.3}$$

If $(x,t) \in \mathcal{M}_t^+$, $r > 0$, and $\Delta t > 0$ then we define the forward *parabolic region* $P(x,t,r,\Delta t)$ to be the union of (the images of) the static admissible curves $\gamma : [t,t'] \to \mathcal{M}$ starting in $B^+(x,t,r)$, where $t' \le t + \Delta t$. That is, we take the union of all the maximal extensions of all static curves, up to time $t + \Delta t$, starting from the initial time slice $B^+(x,t,r)$. When $\Delta t < 0$, the parabolic region $P(x,t,r,\Delta t)$ is defined similarly using static admissible curves ending in $B^-(x,t,r)$.

If $Y \subset \mathcal{M}_t$, and $t \in [c,d]$ then we say that Y is *unscathed in* $[c,d]$ if every point $(x,t) \in Y$ lies on a static curve defined on the time interval $[c,d]$. If, for instance, $d = t$ then this will force $Y \subset \mathcal{M}_t^-$. The term "unscathed" is intended to capture the idea that the set is unaffected by singularities and surgery. (Sometimes Perelman uses the phrase "the solution is defined in $P(x,t,r,\Delta t)$" as synonymous with "the solution is unscathed in $P(x,t,r,\Delta t)$", for example in the definition of canonical neighborhood in II.4.1.) We may use the notation $Y \times [c,d]$ for the set of points lying on static curves $\gamma : [c,d] \to \mathcal{M}$ which pass through Y, when Y is unscathed on $[c,d]$. Note that if Y is open and unscathed on $[c,d]$ then we can think of the Ricci flow on $Y \times [c,d]$ as an ordinary (i.e. surgery-free) Ricci flow.

The definitions of ϵ-neck, ϵ-cap, ϵ-tube and (capped/double) ϵ-horn from Section 58 do not require modification for a Ricci flow with surgery, since they are just special types of Riemannian manifolds; they will turn up as subsets of forward or backward time slices of a Ricci flow with surgery. A *strong ϵ-neck* is a subset of the form $U \times [c,d] \subset \mathcal{M}$, where

$U \subset \mathcal{M}_d^-$ is an open set that is unscathed on the interval $[c, d]$, which is a strong ϵ-neck in the sense of Section 58.

69. II.4.1. *A priori* ASSUMPTIONS

This section introduces the notion of canonical neighborhood.

The following definition captures the geometric structure that emerges by combining Theorem 52.7 and its extension to Ricci flows with surgery (Section 77) with the geometric description of κ-solutions. The idea is that blowups either yield κ-solutions, whose structure is well understood from Section 59, or there are surgeries nearby in the recent past, in which case the local geometry resembles that of the standard solution. Both alternatives produce canonical neighborhoods.

Definition 69.1. (Canonical neighborhoods, cf. Definition in II.4.1) Let $\epsilon > 0$ be small enough so that Lemmas 59.7 and 63.1 hold. Let C_1 be the maximum of $30\epsilon^{-1}$ and the $C_1(\epsilon)$'s of Lemmas 59.7 and 63.1. Let C_2 be the maximum of the $C_2(\epsilon)$'s of Lemmas 59.7 and 63.1. Let $r : [a, b] \to (0, \infty)$ be a positive nonincreasing function. A Ricci flow with surgery $\mathcal{M}$ defined on the time interval $[a, b]$ satisfies the *r-canonical neighborhood assumption* if every $(x, t) \in \mathcal{M}_t^\pm$ with scalar curvature $R(x, t) \geq r(t)^{-2}$ has a canonical neighborhood in the corresponding (forward/backward) time slice, as in Lemma 59.7. More precisely, there is an $\hat{r} \in (R(x, t)^{-\frac{1}{2}}, C_1 R(x, t)^{-\frac{1}{2}})$ and an open set $U \subset \mathcal{M}_t^\pm$ with $\overline{B^\pm(x, t, \hat{r})} \subset U \subset B^\pm(x, t, 2\hat{r})$ that falls into one of the following categories :

(a) $U \times [t - \Delta t, t] \subset \mathcal{M}$ is a strong ϵ-neck for some $\Delta t > 0$. (Note that after parabolic rescaling the scalar curvature at (x, t) becomes 1, so the scale factor must be $\approx R(x, t)$, which implies that $\Delta t \approx R(x, t)^{-1}$.)

(b) U is an ϵ-cap which, after rescaling, is ϵ-close to the corresponding piece of a κ_0-solution or a time slice of a standard solution (cf. Section 60).

(c) U is a closed manifold diffeomorphic to S^3 or RP^3.

(d) U is ϵ-close to a closed manifold of constant positive sectional curvature.

Moreover, the scalar curvature in U lies between $C_2^{-1} R(x, t)$ and $C_2 R(x, t)$. In cases (a), (b), and (c), the volume of U is greater than $C_2^{-1} R(x, t)^{-\frac{3}{2}}$. In case (c), the infimal sectional curvature of U is greater than $C_2^{-1} R(x, t)$.

Finally, we require that

$$|\nabla R(x, t)| < \eta R(x, t)^{\frac{3}{2}}, \qquad \left| \frac{\partial R}{\partial t}(x, t) \right| < \eta R(x, t)^2, \tag{69.2}$$

where η is the constant from (59.5). Here the time dervative $\frac{\partial R}{\partial t}(x, t)$ should be interpreted as a one-sided derivative when the point (x, t) is added or removed during surgery at time t.

Remark 69.3. Note that the smaller of the two balls in $\overline{B^\pm(x, t, \hat{r})} \subset U \subset B^\pm(x, t, 2\hat{r})$ is closed, in order to make it easier to check the openness of the canonical neighborhood condition. The requirement that C_1 be at least $30\epsilon^{-1}$ will be used in the proof of Lemma 73.7; see Remark 73.8.

Remark 69.4. For convenience, in case (b) we have added the extra condition that U is ϵ-close to the corresponding piece of a κ_0-solution or a time slice of a standard solution. One does not need this extra condition, but it is consistent to add it. We remark that when surgery is performed according to the recipe of Section 73, if a point (p,t) lies in $\mathcal{M}_t^+ - \mathcal{M}_t^-$ (i.e. it is "added" by surgery) then it will sit in an ϵ-cap, because $\mathcal{M}_t^+$ will resemble a standard solution from Section 60 near (p,t). Points lying somewhat further out on the capped neck will belong to a strong ϵ-neck which extends backward in time prior to the surgery.

The next condition, which will ultimately be guaranteed by the Hamilton-Ivey curvature pinching result and careful surgery, is also essential in blowup arguments à la Section 52.

Definition 69.5. (Φ-pinching) Let $\Phi \in C^\infty(\mathbb{R})$ be a positive nondecreasing function such that for positive s, $\frac{\Phi(s)}{s}$ is a decreasing function which tends to zero as $s \to \infty$. The Ricci flow with surgery $\mathcal{M}$ satisfies the Φ*-pinching assumption* if for all $(x,t) \in \mathcal{M}$, one has $\mathrm{Rm}(x,t) \geq -\Phi(R(x,t))$.

We remark that the notion of Φ-pinching here is somewhat different from Perelman's ϕ-pinching. The purpose of this definition is to distill out the properties of the Hamilton-Ivey pinching condition which are needed in the rest of the proof.

Definition 69.6. A Ricci flow with surgery satisfies the *a priori assumptions* if it satisfies the Φ-pinching and r-canonical neighborhood assumptions on the time interval of the flow. Note that the *a priori* assumptions depend on ϵ, the function $r(t)$ of Definition 69.1 and the function Φ of Definition 69.5.

70. II.4.2. Curvature bounds from the *a priori* assumptions

In this section we state some technical lemmas about Ricci flows with surgery that satisfy the *a priori* assumptions of the previous section.

The first one is the surgery analog of Lemma 52.11.

Lemma 70.1. *(cf. Claim 1 of II.4.2) Given $(x_0,t_0) \in \mathcal{M}$ put $Q = |R(x_0,t_0)| + r(t_0)^{-2}$. Then $R(x,t) \leq 8Q$ for all $(x,t) \in P(x_0,t_0,\frac{1}{2}\eta^{-1}Q^{-\frac{1}{2}},-\frac{1}{8}\eta^{-1}Q^{-1})$, where η is the constant from (69.2).*

Proof. The lemma follows from the estimates (69.2). One integrates these derivative bounds along a subinterval of a path that goes in $B(x_0,t_0,\frac{1}{2}\eta^{-1}Q^{-\frac{1}{2}})$ and then backward in time along a static path. See the proof of Lemma 52.11. We also use the fact that if $t' \leq t_0$ and $R(x',t') \geq Q$ then the inequalities (69.2) are valid at (x',t'), since $r(\cdot)$ is nonincreasing. □

The next lemma expresses the main consequence of Claim 2 of II.4.2.

Lemma 70.2. *(cf. Claim 2 of II.4.2) If ϵ is small enough then the following holds. Suppose that $\mathcal{M}$ is a Ricci flow with surgery that satisfies the Φ-pinching assumption. Then for any $A < \infty$ and $\widehat{r} > 0$ there exist $\xi = \xi(A) > 0$ and $K = K(A,\widehat{r}) < \infty$ with the following property. Suppose that $\mathcal{M}$ also satisfies the r-canonical neighborhood assumption for some function $r(\cdot)$. Then for any time t_0, if (x_0,t_0) is a point so that $Q = R(x_0,t_0) > 0$ satisfies*

$\frac{\Phi(Q)}{Q} < \xi$ and (x, t_0) is a point so that $\mathrm{dist}_{t_0}(x_0, x) \leq AQ^{-\frac{1}{2}}$ then $R(x, t_0) \leq KQ$, where $K = K(A, r(t_0))$.

Proof. The proof is similar to the proof of Step 2 of Theorem 52.7. (The canonical neighborhood assumption replaces Step 1 of the proof of Theorem 52.7.) Assuming that the lemma fails, one obtains a piece of a nonflat metric cone as a blowup limit. Using the canonical neighborhood assumption, one concludes that the corresponding points in $\mathcal{M}$ have a neighborhood of type (a), i.e. a strong ϵ-neck, since the neighborhoods of type (b), (c) and (d) of Definition 69.1 are not close to a piece of metric cone. A strong ϵ-neck, has the time interval needed to apply the strong maximum principle as in Step 2 of the proof of Theorem 52.7, in order to get a contradiction. □

71. II.4.3. δ-necks in ϵ-horns

In this section we show that an ϵ-horn has a self-improving property as one goes down the horn. For any $\delta > 0$, if the scalar curvature at a point is sufficiently large then the point actually lies in a δ-neck.

In the statement of the next lemma we will write Ω synonymously with the $\mathcal{M}_T^-$ of Section 68.

Lemma 71.1. *(cf. Lemma II.4.3)*

Given the pinching function Φ, a number $\widehat{T} \in (0, \infty)$, a positive nonincreasing function $r : [0, \widehat{T}] \to \mathbb{R}$ and a number $\delta \in \left(0, \frac{1}{2}\right)$, there is a nonincreasing function $h : [0, \widehat{T}] \to \mathbb{R}$ with $0 < h(t) < \delta^2 r(t)$ so that the following property is satisfied. Let $\mathcal{M}$ be a Ricci flow with surgery defined on $[0, T)$, with $T < \widehat{T}$, which satisfies the a priori assumptions (Definition 69.6) and which goes singular at time T. Let $(\Omega, \overline{g})$ denote the time-T limit, in the sense of Section 67. Put $\rho = \delta\, r(T)$ and

$$\Omega_\rho = \{(x, t) \in \Omega \mid \overline{R}(x, T) \leq \rho^{-2}\}. \tag{71.2}$$

Suppose that (x, T) lies in an ϵ-horn $\mathcal{H} \subset \Omega$ whose boundary is contained in Ω_ρ. Suppose also that $\overline{R}(x, T) \geq h^{-2}(T)$. Then the parabolic region $P(x, T, \delta^{-1}\overline{R}(x, T)^{-\frac{1}{2}}, -\overline{R}(x, T)^{-1})$ is contained in a strong δ-neck. (As usual, ϵ is a fixed constant that is small enough so that the result holds uniformly with respect to the other variables.)

Proof. Fix $\delta \in (0, 1)$. Suppose that the claim is not true. Then there is a sequence of Ricci flows with surgery $\mathcal{M}^\alpha$ and points $(x^\alpha, T^\alpha) \in \mathcal{M}^\alpha$ with $T^\alpha < \widehat{T}$ such that
1. $\mathcal{M}^\alpha$ satisfies the Φ-pinching and r-canonical neighborhood assumptions,
2. $\mathcal{M}^\alpha$ goes singular at time T^α,
3. (x^α, T^α) belongs to an ϵ-horn $\mathcal{H}^\alpha \subset \Omega^\alpha$ whose boundary is contained in Ω^α_ρ, and
4. $\overline{R}(x^\alpha, T^\alpha) \to \infty$, but
5. For each α, $P(x^\alpha, T^\alpha, \delta^{-1}\overline{R}(x^\alpha, T^\alpha)^{-\frac{1}{2}}, -\overline{R}(x^\alpha, T^\alpha)^{-1})$ is not contained in a strong δ-neck.

Recall that when ϵ is small enough, any cross-sectional 2-sphere sitting in an ϵ-neck $V \subset \mathcal{H}^\alpha$ separates the ends of $\mathcal{H}^\alpha$; see Section 58. We may find a properly embedded minimizing geodesic $\gamma^\alpha \subset \mathcal{H}^\alpha$ which joins the two ends of $\mathcal{H}^\alpha$. As γ^α must intersect a

cross-sectional 2-sphere containing (x^α, T^α), it must pass within distance $\leq 10\overline{R}(x^\alpha, T^\alpha)^{-\frac{1}{2}}$ of (x^α, T^α), when ϵ is small. Let y^α be the endpoint of γ^α contained in Ω^α_ρ and let $\widehat{y}^\alpha$ be the first point, moving along γ^α from the noncompact end of $\mathcal{H}^\alpha$ toward y^α, where $\overline{R}(\widehat{y}^\alpha, T^\alpha) = \rho^{-2}$. As the gradient bound $|\nabla \overline{R}^{-\frac{1}{2}}| \leq \frac{1}{2}\eta$ is valid along γ^α starting from $\widehat{y}^\alpha$ and going out the noncompact end (since such points on γ^α have scalar curvature greater than $r(T^\alpha)^{-2}$), we have $\mathrm{dist}_{T^\alpha}(x^\alpha, y^\alpha) \geq \mathrm{dist}_{T^\alpha}(x^\alpha, \widehat{y}^\alpha) \geq \frac{2}{\eta}\left(\rho - \overline{R}(x^\alpha, T^\alpha)^{-\frac{1}{2}}\right)$. Let L^α denote the time-T^α distance from x^α to the other end of $\mathcal{H}^\alpha$. Since $\overline{R}$ goes to infinity as one exits the end, Lemma 70.2 implies that $\lim_{\alpha\to\infty} \overline{R}(x^\alpha, T^\alpha)^{\frac{1}{2}} L^\alpha = \infty$. From the existence of γ^α, whose length in either direction from x^α is large compared to $\overline{R}(x^\alpha, T^\alpha)^{-\frac{1}{2}}$, it is clear that for large α, the canonical neighborhood of (x^α, T^α) must be of type (a) or (b) in the terminology of Definition 69.1. By Lemmas 67.9 and 70.2, we also know that for any fixed $\sigma < \infty$, for large α the ball $B(x^\alpha, T^\alpha, \sigma\overline{R}(x^\alpha, T^\alpha)^{-\frac{1}{2}})$ has compact closure in the time-T^α slice of $\mathcal{M}^\alpha$.

By Lemma 70.2, after rescaling the metric on the time-T^α slice by $\overline{R}(x^\alpha, T^\alpha)$ we have uniform curvature bounds on distance balls. We also have a uniform lower bound on the injectivity radius at (x^α, T^α) of the rescaled solution, in view of its canonical neighborhood. Hence after passing to a subsequence, we may take a pointed smooth complete limit $(M^\infty, x^\infty, g_\infty)$ of the time-T^α slices, where the derivative bounds needed to take a smooth limit come from the canonical neighborhood assumption. By the Φ-pinching assumption, M^∞ will have nonnegative curvature.

After passing to a subsequence, we can also assume that the γ^α's converge to a minimizing geodesic γ in M^∞ that passes within distance 10 from x^∞. The rescaled length of γ^α from x^α to y^α is bounded below by $\frac{2}{\eta}\left(\overline{R}(x^\alpha, T^\alpha)^{\frac{1}{2}}\rho - 1\right)$, which tends to infinity as $\alpha \to \infty$. We have shown that the rescaled length of γ^α from x^α to the other end of $\mathcal{H}^\alpha$ also tends to infinity as $\alpha \to \infty$. It follows that γ is bi-infinite. Thus by Toponogov's theorem, M^∞ splits off an $\mathbb{R}$-factor. Then for large α, the canonical neighborhood of (x^α, T^α) must be an ϵ-neck, and $M^\infty = \mathbb{R} \times S^2$ for some positively curved metric on S^2. In particular, M^∞ has scalar curvature uniformly bounded above.

Any point $\widehat{x} \in M^\infty$ is a limit of points $(\widehat{x}^\alpha, T^\alpha) \in \mathcal{M}^\alpha$. As $R_\infty(\widehat{x}) > 0$ and $R(x^\alpha, T^\alpha) \to \infty$, it follows that $R(\widehat{x}^\alpha, T^\alpha) \to \infty$. Then for large α, $(\widehat{x}^\alpha, T^\alpha)$ is in a canonical neighborhood which, in view of the $\mathbb{R}$-factor in M^∞, must be a strong ϵ-neck. From the upper bound on the scalar curvature of M^∞, along with the time interval involved in the definition of a strong ϵ-neck, it follows that we can parabolically rescale the pointed flows $(\mathcal{M}^\alpha, x^\alpha, T^\alpha)$ by $R(x^\alpha, T^\alpha)$, shift time and extract a smooth pointed limiting Ricci flow $(\mathcal{M}^\infty, x^\infty, 0)$ which is defined on a time interval $(\xi, 0]$, for some $\xi < 0$.

In view of the strong ϵ-necks around the points $(\widehat{x}^\alpha, T^\alpha)$, if we take ξ close to zero then we are ensured that the Ricci flow $(\mathcal{M}^\infty, x^\infty, 0)$ has positive scalar curvature R_∞. Given $(\widehat{x}, t) \in \mathcal{M}^\infty$, as $R_\infty(\widehat{x}, t) > 0$ and $R(x^\alpha, T^\alpha) \to \infty$, the Φ-pinching implies that the time-t slice $\mathcal{M}^\infty_t$ has nonnegative curvature at $\widehat{x}$. Thus $\mathcal{M}^\infty$ has nonnegative curvature. The time-0 slice $\mathcal{M}^\infty_0$ splits off an $\mathbb{R}$-factor, which means that the same will be true of all time slices; cf. the proof of Lemma 61.1. Hence $\mathcal{M}^\infty$ is a product Ricci flow.

Let ξ be the minimal negative number so that after parabolically rescaling the pointed flows $(\mathcal{M}^\alpha, x^\alpha, T^\alpha)$ by $R(x^\alpha, T^\alpha)$, we can extract a limit Ricci flow $(\mathcal{M}^\infty, (x^\infty, 0), g_\infty(\cdot))$ which is the product of $\mathbb{R}$ with a positively curved Ricci flow on S^2, and is defined on the time interval $(\xi, 0]$. We claim that $\xi = -\infty$. Suppose not, i.e. $\xi > -\infty$. Given $(\widehat{x}, t) \in \mathcal{M}^\infty$, as $R_\infty(\widehat{x}, t) > 0$ and $R(x^\alpha, T^\alpha) \to \infty$, it follows that (x, t) is a limit of points $(\widehat{x}^\alpha, t^\alpha) \in \mathcal{M}^\alpha$ that lie in canonical neighborhoods. In view of the $\mathbb{R}$-factor in $\mathcal{M}^\infty$, for large α these canonical neighborhoods must be strong ϵ-necks. This implies in particular that $\left(R_\infty^{-2} \frac{\partial R_\infty}{\partial t}\right)(\widehat{x}, t) > 0$, so $\frac{\partial R_\infty}{\partial t}(\widehat{x}, t) > 0$. Then there is a uniform upper bound Q for the scalar curvature on $\mathcal{M}^\infty$. Extending backward from a time-$(\xi + \frac{1}{100Q})$ slice, we can construct a limit Ricci flow that exists on some time interval $(\xi', 0]$ with $\xi' < \xi$. As before, using the strong ϵ-neck condition and the Φ-pinching, if ξ' is sufficiently close to ξ then we are ensured that the Ricci flow on $(\xi', 0]$ is the product of $\mathbb{R}$ with a positively curved Ricci flow on S^2. This is a contradiction.

Thus we obtain an ancient solution $\mathcal{M}^\infty$ with the property that each point (x, t) lies in a strong ϵ-neck. Removing the $\mathbb{R}$-factor gives an ancient solution on S^2. In view of the fact that each time slice is ϵ-close to the round S^2, up to rescaling, it follows that the ancient solution on S^2 must be the standard shrinking solution (see Sections 40 and 43). Then $\mathcal{M}^\infty$ is the standard shrinking solution on $\mathbb{R} \times S^2$. Hence for an infinite number of α, $P(x^\alpha, T^\alpha, \delta^{-1} R(x^\alpha, T^\alpha)^{-\frac{1}{2}}, -R(x^\alpha, T^\alpha)^{-1})$ is in fact in a strong δ-neck, which is a contradiction. $\square$

Remark 71.3. If a given h makes Lemma 71.1 work for a given function r then one can check that logically, h also works for any r' with $r' \geq r$. Because of this, we may assume that h only depends on $\min r = r(T)$ and is monotonically nondecreasing as a function of $r(T)$. Similarly, if a given h makes Lemma 71.1 work for a given value of δ then h also works for any δ' with $\delta' \geq \delta$. Thus we may assume that h is monotonically nondecreasing as a function of δ.

72. Surgery and the pinching condition

This section describes how one can take a δ-neck satisfying the time-t Hamilton-Ivey pinching condition, and perform surgery so as to obtain a new manifold which also satisfies the time-t pinching condition, and which is δ'-close to the standard solution modulo rescaling. Here δ' is a nonexplicit function of δ but satisifes the important property that $\delta'(\delta) \to 0$ as $\delta \to 0$.

The main geometric idea which handles the delicate part of the surgery procedure is contained in the following lemma. It says that one can "round off" the boundary of an approximate round half-cylinder so as to simultaneously increase the scalar curvature and the minimum of sectional curvature at each point.

As the statement of the following lemma involves the curvature operator, we state our conventions. If M has constant sectional curvature k then the curvature operator acts on 2-forms as multiplication by $2k$. This is consistent with the usual Ricci flow literature, e.g. [22].

Recall that ϵ is our global parameter, which is taken sufficiently small.

Lemma 72.1. *Let g_{cyl} denote the round cylindrical metric of scalar curvature 1 on $\mathbb{R} \times S^2$. Let z denote the coordinate in the $\mathbb{R}$-direction. Given $A > 0$, suppose that $f : (-A, 0] \to \mathbb{R}$ is a smooth function such that*

- *$f^{(k)}(0) = 0$ for all $k \geq 0$.*
- *On $(-A, 0)$,*

$$f(z) < 0\,, \quad f'(z) > 0\,, \quad f''(z) < 0. \tag{72.2}$$

- *$\|f\|_{C^2} < \epsilon$.*
- *For every $z \in (-A, 0)$,*

$$\max\left(|f(z)|, |f'(z)|\right) \leq \epsilon\ |f''(z)|\,. \tag{72.3}$$

Then if h_0 is a smooth metric on $(-A, 0] \times S^2$ with $\|h_0 - g_{cyl}\|_{C^2} < \epsilon$ and we set $h_1 = e^{2f(z)} h_0$, it follows that for all $p \in (-A, 0) \times S^2$ we have $R_{h_1}(p) > R_{h_0}(p) - f''(z(p))$. Also, if $\lambda_1(p)$ denotes the lowest eigenvalue of the curvature operator at p then $\lambda_1^{h_1}(p) > \lambda_1^{h_0}(p) - f''(z(p))$.

Proof. We will use the variational characterization of $\lambda_1(p)$:

$$\lambda_1(p) \;=\; \inf_{\omega \neq 0} \frac{\omega_{ij}\, R^{ij}{}_{kl}\, \omega^{kl}}{\omega_{ij}\, \omega^{ij}} \tag{72.4}$$

where $\omega \in \Lambda^2(T_pM)$. We will also use the following formulas about curvature quantities for conformally related metrics in dimension 3 :

$$R_{h_1} \;=\; e^{-2f}\ \left(R_{h_0} \;-\; 4 \triangle f \;-\; 2\, |\nabla f|^2\right) \tag{72.5}$$

and

(72.6)

$$R^{ij}{}_{kl}(h_1) = e^{-2f} \left(R^{ij}{}_{kl}(h_0) \;-\; \widetilde{f}^i{}_k\, \delta^j{}_l + \widetilde{f}^i{}_l\, \delta^j{}_k + \widetilde{f}^j{}_k\, \delta^i{}_l - \widetilde{f}^j{}_l\, \delta^i{}_k \;-\; |\nabla f|^2 \left(\delta^i{}_k\, \delta^j{}_l \;-\; \delta^i{}_l\, \delta^j{}_k\right)\right),$$

where $\widetilde{f}_{ij} \;=\; f_{;ij} \;-\; f_{;i}\, f_{;j}$. That is, $\widetilde{f} \;=\; \mathrm{Hess}(f) \;-\; df \,\otimes\, df$. The right-hand sides of these expressions are computed using the metric h_0.

To motivate the proof, let us first consider the linearization of these expressions around h_0. Keeping only the linear terms in f gives to leading order,

$$R_{h_1} \;\sim\; R_{h_0} - 2f\, R_{h_0} \;-\; 4 \triangle f \tag{72.7}$$

and

$$R^{ij}{}_{kl}(h_1) \;\sim\; R^{ij}{}_{kl}(h_0) \;-\; 2f\, R^{ij}{}_{kl}(h_0) \;-\; f_{;}{}^i{}_k\, \delta^j{}_l \;+\; f_{;}{}^i{}_l\, \delta^j{}_k \;+\; f_{;}{}^j{}_k\, \delta^i{}_l \;-\; f_{;}{}^j{}_l\, \delta^i{}_k. \tag{72.8}$$

From the assumptions, $R_{h_0} \sim 1$ and $f < 0$ on $(-A, 0) \times S^2$. As h_0 is close to g_{cyl}, we have $\triangle f \sim f''(z)$, so $-2f\, R_{h_0} \;-\; 4 \triangle f \;\geq\; -\, f''(z)$. Similarly, in the case of g_{cyl} a minimizer ω in (72.4) is of the form $\omega \;=\; X \wedge \partial_z$, where X is a unit vector in the S^2-direction. As h_0 is close to g_{cyl}, a minimizing ω for h_0 will be close to something of the form $X \wedge \partial_z$. Then

$$\lambda_1^{h_1} \sim \lambda_1^{h_0} \;-\; 2f\, \lambda_1^{h_0} \;-\; 2\, f''(z) \;\geq\; \lambda_1^{h_0} \;+\; 2f(z)\, |\lambda_1^{h_0}| \;-\; 2\, f''(z). \tag{72.9}$$

As h_0 is close to g_{cyl}, $\lambda_1^{h_0}$ is close to $\lambda_1^{g_{cyl}} = 0$. Then we can use (72.3) to say that $2f(z)|\lambda_1^{h_0}| - 2\, f''(z) \;\geq\; -\, f''(z)$.

The remaining issue is to show that the increase in R and λ_1 coming from the linear approximation is still approximately valid in the nonlinear case, provided that ϵ is sufficiently small. For this, we have to show that the increase from the linear approximation dominates the error terms that we have neglected.

To deal with the scalar curvature first, from (72.5) we have

$$\begin{aligned} R_{h_1} &= e^{-2f}\left(R_{h_0} - 4\,\triangle_{g_{cyl}} f\right) + 4e^{-2f}\left(\triangle_{g_{cyl}} f - \triangle f\right) - 2\,e^{-2f}\,|\nabla f|^2 \\ &\geq R_{h_0} - 4\,f''(z) + 4e^{-2f}\left(\triangle_{g_{cyl}} f - \triangle f\right) - 2\,e^{-2f}\,|\nabla f|^2. \end{aligned} \tag{72.10}$$

Next, there is an estimate of the form

$$\begin{aligned} |\triangle_{g_{cyl}} f - \triangle f| &\leq \text{const.}\, \| h_0 - g_{cyl} \|_{C^2}\, \left(|f(z)| + |f'(z)| + |f''(z)|\right) \\ &\leq \text{const.}\, \epsilon\, \left(|f(z)| + |f'(z)| + |f''(z)|\right). \end{aligned} \tag{72.11}$$

As $e^{-2f} \leq e^{2\epsilon}$, if ϵ is small then

$$\left|e^{-2f}\left(\triangle_{g_{cyl}} f - \triangle f\right)\right| \leq \text{const.}\, \epsilon\, \left(|f(z)| + |f'(z)| + |f''(z)|\right) \tag{72.12}$$

Similarly,

$$e^{-2f}\,|\nabla f|^2 \leq \text{const.}\, |f'(z)|^2 \leq \text{const.}\, \epsilon\, |f'(z)|. \tag{72.13}$$

When combined with (72.3), if ϵ is taken sufficiently small then

$$-4\,f''(z) + 4e^{-2f}\left(\triangle_{g_{cyl}} f - \triangle f\right) - 2\,e^{-2f}\,|\nabla f|^2 \geq -f''(z). \tag{72.14}$$

This shows the desired estimate for $R_{h_1}(p)$.

To estimate $\lambda_1^{h_1}$ we use (72.4) and (72.6) to write

$$\lambda_1^{h_1}(p) = e^{-2f(z)}\left(\inf_{\omega\neq 0} \frac{\omega_{ij}\, R^{ij}{}_{kl}(h_0)\,\omega^{kl} - 4\,\omega_{ij}\widetilde{f}^i{}_k\omega^{kj}}{\omega_{ij}\,\omega^{ij}} - 2\,|\nabla f|^2(z)\right). \tag{72.15}$$

Comparing with

$$\lambda_1^{h_0}(p) = \inf_{\omega\neq 0} \frac{\omega_{ij}\, R^{ij}{}_{kl}(h_0)\,\omega^{kl}}{\omega_{ij}\,\omega^{ij}} \tag{72.16}$$

gives

$$\lambda_1^{h_0}(p) \leq e^{2f(z)}\,\lambda_1^{h_1}(p) + \frac{4\,\omega_{ij}\widetilde{f}^i{}_k\omega^{kj}}{\omega_{ij}\,\omega^{ij}} + 2\,|\nabla f|^2(z), \tag{72.17}$$

where ω is a minimizer in (72.15), or

$$\lambda_1^{h_1}(p) \geq e^{-2f(z)}\,\lambda_1^{h_0}(p) - 4\,e^{-2f(z)}\,\frac{\omega_{ij}\widetilde{f}^i{}_k\omega^{kj}}{\omega_{ij}\,\omega^{ij}} - 2\,e^{-2f(z)}\,|\nabla f|^2(z). \tag{72.18}$$

Using the variational formula (72.4), one can show that $|\lambda_1^{h_0}(p)| \leq \text{const.}\,\epsilon$. From eigenvalue perturbation theory [55, Chapter 12], ω will be of the form $X \wedge \partial_z + O(\epsilon)$ for some unit vector X tangential to S^2. Then we get an estimate

$$\lambda_1^{h_1}(p) \geq \lambda_1^{h_0}(p) - 2f''(z) - \text{const.}\, \epsilon\, \left(|f(z)| + |f'(z)| + |f''(z)|\right). \tag{72.19}$$

From (72.3), if ϵ is taken sufficiently small then $\lambda_1^{h_1}(p) - \lambda_1^{h_0}(p) \geq -f''(z)$. $\square$

Recall that the initial condition $\mathcal{S}_0$ for the standard solution is an $O(3)$-symmetric metric g_0 on $\mathbb{R}^3$ with nonnegative curvature operator, whose end is isometric to a round half-cylinder of scalar curvature 1. To facilitate the surgery procedure, we will assume that some metric ball around the $O(3)$-fixed point has constant positive curvature. Outside of this ball we use radial coordinates $(z, \theta) \in (-B, \infty) \times S^2$, with $g_0 = e^{2F(z)} g_{cyl}$. Here g_{cyl} is the round cylindrical metric of scalar curvature one and $F \in C^\infty(-B, \infty)$.

Lemma 72.20. *Given $A > 0$, we can choose $B > A$ and $F \in C^\infty(-B, \infty)$ so that*
1. $F \equiv 0$ on $[0, \infty) \times S^2$.
2. The restriction of F to $(-A, 0] \times S^2$ satisfies the hypotheses of Lemma 72.1.
3. The metric $e^{2F(z)} g_{cyl}$ on $(-B, \infty) \times S^2$ has nonnegative sectional curvature and extends smoothly to a metric on $\mathbb{R}^3$ by adding a ball of constant positive curvature at $\{-B\} \times S^2$.

Proof. For a metric of the form $e^{2F(z)} g_{cyl}$, one computes that the sectional curvatures are $-\, e^{-\,2F}\, F''$ and $e^{-\,2F} \left(\frac{1}{2} \,-\, (F')^2\right)$. In particular, the conditions for positive sectional curvature are $F'' < 0$ and $|F'| < \frac{1}{\sqrt{2}}$.

The 3-sphere of constant sectional curvature k^2, with two points removed, has a metric given by

$$F_k(z) \;=\; \log\left(\frac{\sqrt{2}}{k}\right) \;+\; \frac{1}{\sqrt{2}}\, z \;-\; \log\left(1 + e^{\sqrt{2} z}\right). \tag{72.21}$$

(Shifting z gives other metrics of constant curvature k^2. We have normalized so that the $z = 0$ slice is the slice of maximal area.) Note that the derivative

$$D(z) \;=\; \frac{1}{\sqrt{2}} \;-\; \sqrt{2}\, \frac{e^{\sqrt{2} z}}{1 + e^{\sqrt{2} z}} \tag{72.22}$$

is independent of k.

Given $A > 0$, we take F to be 0 on $[0, \infty)$ and of the form $c_1\, e^{c_2/z}$ on $(-A, 0]$. We can take the constant $c_1 > 0$ sufficiently small and the constant $c_2 < \infty$ sufficiently large so that the hypotheses of Lemma 72.1 are satisfied. It remains to smoothly cap off $([-A, \infty) \times S^2, e^{2F(z)} g_{cyl})$ with something of positive sectional curvature.

With our given choice of $F\big|_{(-A,0]}$, we have $F'(-A) \in (0, \epsilon)$. As $\lim_{z \to -\infty} D(z) \;=\; \frac{1}{\sqrt{2}}$, we can choose $B > A$ so that $D(-B) > F'(-A)$. As $F''(-A) < 0$ and $D'(-B) < 0$, we can extend F' to a smooth function $\widetilde{D} \,:\, (-B, \infty) \to \left(0, \frac{1}{\sqrt{2}}\right)$ which has $\widetilde{D}' < 0$ and which coincides with D on a small interval $(-B, -B + \delta)$. Putting

$$F(z) \;=\; F(0) \;+\; \int_0^z \widetilde{D}(w)\, dw, \tag{72.23}$$

we obtain $F \in C^\infty(B, \infty)$ which coincides with F_k on $(-B, -B + \delta)$, for some $k > 0$. Then we can glue on a round metric ball of constant curvature k^2 to $\{-B\} \times S^2$, in order to obtain the desired metric. $\square$

In the statement of the next lemma we continue with the metric constructed in Lemma 72.20.

Lemma 72.24. *There exists $\delta' = \delta'(\delta)$ with $\lim_{\delta \to 0} \delta'(\delta) = 0$ and a constant $\delta_0 > 0$ such that the following holds. Suppose that $\delta < \delta_0$, $x \in \{0\} \times S^2$ and h_0 is a Riemannian metric on $(-A, \frac{1}{\delta}) \times S^2$ with $R(x) > 0$ such that:*

- *h_0 satisfies the time-t Hamilton-Ivey pinching condition of Definition B.5.*
- *$R(x)h_0$ is δ-close to g_{cyl} in the $C^{[\frac{1}{\delta}]+1}$-topology.*

Then there is a smooth metric h on $\mathbb{R}^3 = D^3 \cup \left((-B, \frac{1}{\delta}) \times S^2\right)$ such that

- *h satisfies the time-t pinching condition.*
- *The restriction of h to $[0, \frac{1}{\delta}) \times S^2$ is h_0.*
- *The restriction of $R(x)h$ to $(-B, -A) \times S^2$ is g_0, the initial metric of a standard solution.*
- *The restriction of $R(x)h$ to D^3 has constant curvature k^2.*
- *$R(x)h$ is δ'-close to $e^{2F} g_{cyl}$ in the $C^{[\frac{1}{\delta'}]+1}$-topology on $\left(-B, \frac{1}{\delta}\right) \times S^2$.*

Proof. Put

$$U_1 = (-B, -\frac{A}{2}) \times S^2, \quad U_2 = (-A, \frac{1}{\delta}) \times S^2 \tag{72.25}$$

and let $\{\alpha_1, \alpha_2\}$ be a C^∞ partition of unity subordinate to the open cover $\{U_1, U_2\}$ of $(-B, \frac{1}{\delta}) \times S^2$. We set

$$h = \alpha_1 \, R(x)^{-1} \, g_0 + \alpha_2 \, e^{2F} \, h_0 \tag{72.26}$$

on $\left(-B, \frac{1}{\delta}\right) \times S^2$ and cap it off with a 3-ball of constant curvature k, as in Lemma 72.20.

Given δ', we claim that if δ is sufficiently small then the conclusion of the lemma holds. The only part of the lemma that is not obvious is the pinching condition. Note that on $\left(-\frac{A}{2}, \frac{1}{\delta}\right) \times S^2$ the metric h agrees with $e^{2F} h_0$ and hence, when δ is sufficiently small, the pinching condition will hold on $\left(-\frac{A}{2}, \frac{1}{\delta}\right) \times S^2$ by Lemmas 72.1 and B.6. On the other hand, when δ is sufficiently small, the restrictions of the metrics $g_0 = e^{2F} g_{cyl}$ and $R(x)e^{2F} h_0$ to $(-A, -\frac{A}{2}) \times S^2$ will be very close and will have strictly positive curvature. (The positive curvature for $e^{2F} h_0$ also follows from Lemma 72.1; if δ is small enough then $\lambda_1^{h_0}$ will be close to zero, while $-f''(z)$ is strictly positive for $z \in (-A, -\frac{A}{2})$.) Thus h will have positive curvature on $(-B, -\frac{A}{2}) \times S^2$ and the pinching condition will hold there. □

We have now fixed the initial condition g_0 for a standard solution, along with the procedure to meld g_0 to an approximate cylinder.

73. II.4.4. Performing surgery and continuing flows

This section discusses the surgery procedure and shows how to prolong a Ricci flow with surgery, provided that the *a priori* assumptions hold.

Definition 73.1. (Ricci flow with cutoff) Suppose that $a \geq 0$ and let $\mathcal{M}$ be a Ricci flow with surgery defined on $[a, b]$ that satisfies the *a priori* assumptions of Definition 69.6. Let $\delta : [a, b] \to (0, \delta_0)$ be a nonincreasing function, where δ_0 is the parameter of Lemma 72.24. Then $\mathcal{M}$ is a *Ricci flow with* (r, δ)*-cutoff* if at each singular time t, the forward time slice $\mathcal{M}_t^+$ is obtained from the backward time slice $\Omega = \mathcal{M}_t^-$ by applying the following procedure:

A. Discard each component of Ω that does not intersect

$$\Omega_\rho = \{(x,t) \in \Omega \mid R(x,t) \leq \rho^{-2}\}, \tag{73.2}$$

where $\rho = \delta(t) r(t)$.

B. In each ϵ-horn $\mathcal{H}_{ij}$ of each of the remaining components Ω_i, find a point (x_{ij}, t) such that $R(x_{ij}, t) = h^{-2}$, where $h = h(t)$ is as in Lemma 71.1.

C. Find a strong δ-neck $U_{ij} \times [t - h^2, t]$ containing $P(x_{ij}, t, \delta^{-1} R(x_{ij}, t)^{-\frac{1}{2}}, -R(x_{ij}, t)^{-1})$; this is guaranteed to exist by Lemma 71.1.

D. For each ij, let $S_{ij} \subset U_{ij}$ be a cross-sectional 2-sphere containing (x_{ij}, t). Cut $\bigcup_i \Omega_i$ along the S_{ij}'s and throw away the tips of the horns $\mathcal{H}_{ij}$, to obtain a compact manifold-with-boundary X having a spherical boundary component for each ij.

E. Glue caps onto X, using Lemma 72.24, to obtain the closed manifold $\mathcal{M}_t^+$.

For concreteness, we take the parameter A of Lemma 72.24 to be 10. The neighborhood of a boundary component of X is parametrized as $[-A, \delta^{-1}) \times S^2$, with $S_{ij} = \{-A\} \times S^2$. The metric on $[0, \delta^{-1}) \times S^2$ is unaltered by the surgery procedure. The corresponding region in the new manifold $\mathcal{M}_t^+$, minus a metric ball of constant curvature, is parametrized by $(-B, \delta^{-1}) \times S^2$. Put $S'_{ij} = \{0\} \times S^2 \subset \mathcal{M}_t^-$. We will consider the part added by surgery on $\mathcal{H}_{ij}$ to be the 3-disk in $\mathcal{M}_t^+$ bounded by S'_{ij}. In terms of Definition 68.1, if $t = t_k^+$ then the subset X_k^+ of $\Omega_k = \mathcal{M}_t^-$ has boundary $\bigcup_{ij} S'_{ij}$. The added part $\mathcal{M}_t^+ - X_{k+1}^-$ is a union of 3-balls.

Remark 73.3. Our definition of surgery differs slightly from that in [52]. The paper [52] has two extra steps involving throwing away certain components of the postsurgery manifold. We omit these steps in order to simplify the definition of surgery, but there is no real loss either way.

First, in the setup of [52, Section 4.4], any component of $\mathcal{M}_t^+$ that is ϵ-close to a metric quotient of the round S^3 is thrown away. The motivation of [52] was to not have to include these in the list of canonical neighborhoods. Such components are topologically standard. We do include such manifolds in the list of canonical neighborhoods and do not throw them away in the surgery procedure.

Second, when considering the long-time behavior of Ricci flow in [52, Section 7], any component of $\mathcal{M}_t^+$ which admits a metric of nonnegative scalar curvature is thrown away. The motivation for this extra step is that any such component admits a metric that is either flat or has finite extinction time. In either case one concludes that the component is a graph manifold and, for the purposes of the geometrization conjecture, is standard. (Recall the definition of graph manifolds from Appendix I.) Again, we do not throw away such components.

Note that the definition of Ricci flow with (r, δ)-cutoff also depends on the function $r(t)$ through the *a priori* assumption. We now state how the topology of the time slice changes when going backward through the singular time t. Recall that for $t' < t$ close to t, the time slices $\mathcal{M}_{t'}$ are all diffeomorphic; we refer to this diffeomorphism type as the *presurgery manifold*, and the forward time slice $\mathcal{M}_t^+$ as the *postsurgery manifold*.

Lemma 73.4. *The presurgery manifold may be obtained from the postsurgery manifold by applying the following operations finitely many times:*

- *Replacing two connected components with their connected sum.*
- *Taking connected sum of a connected component with $S^1 \times S^2$ or $\mathbb{R}P^3$.*
- *Taking the disjoint union with an additional $S^1 \times S^2$ or an isometric quotient of the round S^3.*

Proof. The proof is basically the same as that of Lemma 67.13. The only difference is that we must take into account the compact components of Ω that do not intersect Ω_ρ; these are thrown away in Step A. (Such components did not occur in Lemma 67.13 because in Lemma 67.13 we were dealing with the first surgery for the Ricci flow on the initial connected manifold; see Lemma 67.4, which is valid for the first surgery time.) Any such component is diffeomorphic to $S^1 \times S^2$, $\mathbb{R}P^3 \# \mathbb{R}P^3$ or a quotient of the round S^3, in view of the canonical neighborhood assumption; see the proof of Lemma 67.5. □

Remark 73.5. When $\delta > 0$ is sufficiently small, we will have $\mathrm{vol}(\mathcal{M}_t^+) < \mathrm{vol}(\mathcal{M}_t^-) - h(t)^3$ for each surgery time $t \in (a, b)$. This is because each component that is discarded in step D contains at least "half" of the δ-neck U_{ij}, which has volume at least const. $\delta^{-1} h(t)^3$, while the cap added has volume at most const. $h(t)^3$.

Remark 73.6. For a Ricci flow with surgery whose original manifold is nonaspherical and irreducible, one wants to know that the Ricci flow goes extinct within a finite time [24, 25, 53]. Consider the effect of a first surgery, say at time t. Among the connected components of the postsurgery manifold $\mathcal{M}_t^+$, one will be diffeomorphic to the presurgery manifold and the others will be 3-spheres. Let $\mathcal{N}_t^+$ be a component of $\mathcal{M}_t^+$ that is diffeomorphic to the presurgery manifold. By the nature of the surgery procedure, there is a function ξ defined on a small interval $(t - \alpha, t)$ so that $\lim_{t' \to t} \xi(t') = 1$ and for $t' \in (t - \alpha, t)$, there is a homotopy-equivalence from $(\mathcal{M}_{t'}, g(t'))$ to $\mathcal{N}_t^+$ that expands distances by at most $\xi(t')$. Following the subsequent evolution of $\mathcal{N}_t^+$, there is a similar statement for the later singular times. This fact is needed in [24, 25, 53] in order to control the decay of a certain area functional as one goes through a surgery.

We discuss how to continue Ricci flows after surgery. We recall that in Definition 68.1 of a Ricci flow with surgery defined on an interval $[a, c]$, the final time slice $\mathcal{M}_c$ consists of a single manifold $\mathcal{M}_c^- = \Omega$ that may or may not be singular.

Lemma 73.7. *(Prolongation of Ricci flows with cutoff) Take the function Φ to be the time-dependent pinching function associated to Definition B.5 in Appendix B. Suppose that r and δ are nonincreasing positive functions defined on $[a, b]$. Let $\mathcal{M}$ be a Ricci flow with (r, δ)-cutoff defined on an interval $[a, c] \subset [a, b]$. Provided* $\sup \delta$ *is sufficiently small, either*

(1) *$\mathcal{M}$ can be prolonged to a Ricci flow with (r, δ)-cutoff defined on $[a, b]$, or*

(2) *There is an extension of $\mathcal{M}$ to a Ricci flow with surgery defined on an interval $[a, T]$ with $T \in (c, b]$, where*
a. The restriction of the flow to any subinterval $[a, T']$, $T' < T$, is a Ricci flow with (r, δ)-cutoff, but
b. The r-canonical neighborhood assumption fails at some point $(x, T) \in \mathcal{M}_T^-$.

In particular, the only obstacle to prolongation of Ricci flows with (r, δ)-cutoff is the potential breakdown of the r-canonical neighborhood assumption.

Proof. Consider the time slice of $\mathcal{M}_c = \mathcal{M}_c^-$ at time c. If it is singular then we perform steps A-E of Definition 73.1 to produce $\mathcal{M}_c^+$; otherwise we set $\mathcal{M}_c^+ = \mathcal{M}_c^-$. Since the surgery is done using Lemma 72.24, provided $\delta > 0$ is sufficiently small, the forward time slice $\mathcal{M}_c^+$ will satisfy the Φ-pinching assumption.

We claim that the r-canonical neighborhood assumption holds in $\mathcal{M}_c^+$. More precisely, if a point $(x, c) \in \mathcal{M}_c^+$ lies within a distance of $10\epsilon^{-1}h$ from the added part $\mathcal{M}_c^+ - \mathcal{M}_c^-$ then it lies in an ϵ-cap, while if (x, c) lies at distance greater than $10\epsilon^{-1}h$ from $\mathcal{M}_c^+ - \mathcal{M}_c^-$ and has scalar curvature greater than $r(c)^{-2}$ then it lies in a canonical neighborhood that was present in the presurgery manifold $\mathcal{M}_c^-$. (We are assuming that $\epsilon < \frac{1}{100}$.) In view of Lemma 63.1, the only point to observe is that points at distance roughly $10\epsilon^{-1}h$ lie in ϵ-necks, as they are unaltered by the surgery and they were in δ-necks before the surgery. This gives the ϵ-neck needed to define an ϵ-cap.

We now prolong $\mathcal{M}$ by Ricci flow with initial condition $\mathcal{M}_c^+$. If the flow extends smoothly up to time b then we are done because either the canonical neighborhood assumption holds up to time b yielding (a), or it fails at some time in the interval $(c, b]$, and we have (b). Otherwise, there is some time $t_{\text{sing}} \le b$ at which it goes singular. We add the singular limit Ω at time t_{sing} to obtain a Ricci flow with surgery defined on $[a, t_{\text{sing}}]$. From Lemma 72.24, $\mathcal{M}$ satisfies the Hamilton-Ivey pinching condition of Definition B.5 on $[a, t_{\text{sing}}]$. As the function r is nonincreasing in t, it follows from Definition 69.1 that the set of times $t \in [c, t_{\text{sing}}]$ for which the r-canonical neighborhood assumption holds is relatively open to the right (i.e. if the r-canonical neighborhood assumption holds at time $t \in [c, t_{\text{sing}})$ then it also holds within some interval $[t, t')$). Thus the set of times $t \in [c, t_{\text{sing}}]$ for which the r-canonical neighborhood assumption holds is either an interval $[c, T)$, with $T \le t_{\text{sing}}$, or $[c, t_{\text{sing}}]$. If the set of such times t is (c, T) for some $T \le t_{\text{sing}}$ then the lemma holds. Otherwise, the r-canonical neighborhood assumption holds at t_{sing}. In this case we repeat the construction with c replaced by t_{sing}, and iterate if necessary. Either we will reach time b after a finite number of iterations, or we will reach a time T satisfying (2), or we will hit an infinite number of singular times before time b. However, the last possibility cannot occur. A singular time corresponds to a component going extinct or to a surgery. The number of components going extinct before time b can be bounded in terms of the number of surgeries before time b, so it suffices to show that the latter is finite. Each surgery removes a volume of at least h^3, but the lower bound on the scalar curvature during the flow, coming from the maximum principle, gives a finite upper bound on the total volume growth during the complement of the singular times. $\square$

Remark 73.8. The condition $C_1 \ge 30\epsilon^{-1}$ in Definition 69.1 was in order to ensure that the ϵ-cap coming from a surgery satisfies the requirements to be a canonical neighborhood.

74. II.4.5. Evolution of a surgery cap

Let $\mathcal{M}$ be a Ricci flow with (r, δ)-cutoff. The next result says that provided δ is small, after a surgery at scale h there is a ball B of radius $Ah \gg h$ centered in the surgery cap whose evolution is close to that of a standard solution for an elapsed time close to h^2, unless another surgery occurs during which the entire ball is thrown away. Note that the elapsed time h^2 corresponds, modulo parabolic rescaling, to the duration of the standard solution.

Lemma 74.1. *(cf. Lemma II.4.5)*
For any $A < \infty$, $\theta \in (0,1)$ and $\hat{r} > 0$, one can find $\hat{\delta} = \hat{\delta}(A, \theta, \hat{r}) > 0$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff defined on a time interval $[a, b]$ with $\min r = r(b) \geq \hat{r}$. Suppose that there is a surgery time $T_0 \in (a, b)$, with $\delta(T_0) \leq \hat{\delta}$. Consider a given surgery at the surgery time and let $(p, T_0) \in \mathcal{M}^+_{T_0}$ be the center of the surgery cap. Let $\hat{h} = h(\delta(T_0), \epsilon, r(T_0), \Phi)$ be the surgery scale given by Lemma 71.1 and put $T_1 = \min(b, T_0 + \theta \hat{h}^2)$. Then one of the two following possibilities occurs :
(1) The solution is unscathed on $P(p, T_0, A\hat{h}, T_1 - T_0)$. The pointed solution there (with respect to the basepoint (p, T_0)) is, modulo parabolic rescaling, A^{-1}-close to the pointed flow on $U_0 \times [0, (T_1 - T_0)\hat{h}^{-2}]$, where U_0 is an open subset of the initial time slice $\mathcal{S}_0$ of a standard solution $\mathcal{S}$ and the basepoint is the center c of the cap in $\mathcal{S}_0$.
(2) Assertion (1) holds with T_1 replaced by some $t^+ \in [T_0, T_1)$, where t^+ is a surgery time. Moreover, the entire ball $B(p, T_0, A\hat{h})$ becomes extinct at time t^+, i.e. $P(p, T_0, A\hat{h}, t_+ - T_0) \cap \mathcal{M}_{t_+} \subset \mathcal{M}^-_{t_+} - \mathcal{M}^+_{t_+}$.

Proof. We give a proof with the same ingredients as the proof in [52], but which is slightly rearranged. We first show the following result, which is almost the same as Lemma 74.1.

Lemma 74.2. *For any $A < \infty$, $\theta \in (0,1)$ and $\hat{r} > 0$, one can find $\hat{\delta} = \hat{\delta}(A, \theta, \hat{r}) > 0$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff defined on a time interval $[a, b]$ with $\min r = r(b) \geq \hat{r}$. Suppose that there is a surgery time $T_0 \in (a, b)$, with $\delta(T_0) \leq \hat{\delta}$. Consider a given surgery at the surgery time and let $(p, T_0) \in \mathcal{M}^+_{T_0}$ be the center of the surgery cap. Let $\hat{h} = h(\delta(T_0), \epsilon, r(T_0), \Phi)$ be the surgery scale given by Lemma 71.1 and put $T_1 = \min(b, T_0 + \theta \hat{h}^2)$. Suppose that the solution is unscathed on $P(p, T_0, A\hat{h}, T_1 - T_0)$. Then the pointed solution there (with respect to the basepoint (p, T_0)) is, modulo parabolic rescaling, A^{-1}-close to the pointed flow on $U_0 \times [0, (T_1 - T_0)\hat{h}^{-2}]$, where U_0 is an open subset of the initial time slice $\mathcal{S}_0$ of a standard solution $\mathcal{S}$ and the basepoint is the center c of the cap in $\mathcal{S}_0$.*

Proof. Fix θ and $\hat{r}$. Suppose that the lemma is not true. Then for some $A > 0$, there is a sequence $\{\mathcal{M}^\alpha, (p^\alpha, T_0^\alpha)\}_{\alpha=1}^\infty$ of pointed Ricci flows with $(r^\alpha, \delta^\alpha)$-cutoff that together provide a counterexample. In particular,
1. $\lim_{\alpha \to \infty} \delta^\alpha(T_0^\alpha) = 0$.
2. $\mathcal{M}^\alpha$ is unscathed on $P(p^\alpha, T_0^\alpha, A\hat{h}^\alpha, T_1^\alpha - T_0^\alpha)$.
3. If $(\widehat{\mathcal{M}}^\alpha, (\hat{p}^\alpha, 0))$ is the pointed Ricci flow arising from $(\mathcal{M}^\alpha, (p^\alpha, T_0^\alpha))$ by a time shift of T_0^α and a parabolic rescaling by $\hat{h}^\alpha$ then $P(\hat{p}^\alpha, 0, A, (T_1^\alpha - T_0^\alpha)(\hat{h}^\alpha)^{-2})$ is not A^{-1}-close to a pointed subset of a standard solution.

Put $T_2 = \liminf_{\alpha\to\infty}(T_1^\alpha - T_0^\alpha)(\hat{h}^\alpha)^{-2}$. (We do not exclude that $T_2 = 0$.) Clearly $T_2 \le \theta$. After passing to a subsequence, we can assume that $T_2 = \lim_{\alpha\to\infty}(T_1^\alpha - T_0^\alpha)(\hat{h}^\alpha)^{-2}$. Let T_3 be the supremum of the set of times $\tau \in [0, T_2]$ with the property that we can apply Appendix E, if we want, to take a convergent subsequence of the pointed solutions $(\widehat{\mathcal{M}}^\alpha, (\hat{p}^\alpha, 0))$ on the time interval $[0, \tau]$ to get a limit solution with bounded curvature. (In applying Appendix E, we use the case $l > 0$ of Appendix D to get bounds on the curvature derivatives near time 0. In particular, if $T_2 > 0$ then $T_3 > 0$.) From the nature of the surgery gluing in Lemma 72.24, since $\lim_{\alpha\to\infty}\delta^\alpha(T_0^\alpha) = 0$ we know that we can at least take a limit of the pointed solutions $(\widehat{\mathcal{M}}^\alpha, (\hat{p}^\alpha, 0))$ on the time interval $[0, 0]$, so T_3 is well-defined.

Sublemma 74.3. *$T_3 = T_2$.*

Proof. Suppose not. Consider the interval $[0, T_3)$ (where we define $[0, 0)$ to be $\{0\}$). Given $\sigma \in (0, T_2 - T_3]$, for any subsequence of $\{\mathcal{M}^\alpha, (\hat{p}^\alpha, 0)\}_{\alpha=1}^\infty$ (which we relabel as $\{\mathcal{M}^\alpha, (\hat{p}^\alpha, 0)\}_{\alpha=1}^\infty$) either

1. There is some $\lambda > 0$ and an infinite number of α for which the set $B(\hat{p}^\alpha, 0, \lambda)$ becomes scathed on $[0, T_3 + \sigma]$, or
2. For each $\lambda > 0$ the set $P(\hat{p}^\alpha, 0, \lambda, T_3 + \sigma)$ is unscathed for large α, but for each $\Lambda > 0$ there is some $\lambda_\Lambda > 0$ such that $\limsup_{\alpha\to\infty}\sup_{P(\hat{p}^\alpha, 0, \lambda_\Lambda, T_3+\sigma)} |\operatorname{Rm}| \ge \Lambda$.

By Appendix E, after passing to a subsequence, there is a complete limit solution $(\widehat{\mathcal{M}}^\infty, (\hat{p}^\infty, 0))$ defined on the time interval $[0, T_3)$ with bounded curvature on compact time intervals. Relabel the subsequence by α. By Lemma 60.3, $(\widehat{\mathcal{M}}^\infty, (\hat{p}^\infty, 0))$ must be the same as the restriction of some standard solution to $[0, T_3)$. From Lemma 62.1, the curvature of $\widehat{\mathcal{M}}^\infty$ is uniformly bounded on $[0, T_3)$; therefore by the canonical neighborhood assumption and equation (69.2), we can choose $\sigma \in (0, T_2 - T_3]$ and $\Lambda' > 0$ so that for any $\lambda > 0$, we have $\limsup_{\alpha\to\infty}\sup_{P(\hat{p}^\alpha, 0, \lambda, T_3+\sigma)} |\operatorname{Rm}| \le \Lambda'$. However, $\lim_{\alpha\to\infty}\delta^\alpha(T_0^\alpha) = 0$ and surgeries only occur near the centers of δ-necks. From the curvature bound on the time interval $[0, T_3 + \sigma]$ and the length distortion estimates of Lemma 27.8, for a given λ the balls $B(\hat{p}^\alpha, 0, \lambda)$ will stay within a uniformly bounded distance from $\hat{p}^\alpha$ on the time interval $[0, T_3 + \sigma]$. Hence they cannot be scathed on $[0, T_3 + \sigma]$ for an infinite number of α, as the collar length of the δ-neck around the supposed surgery locus would be large enough to prohibit the cap point $\hat{p}^\alpha$ from being within a bounded distance from the surgery locus. This, along with the fact that $\limsup_{\alpha\to\infty}\sup_{P(\hat{p}^\alpha, 0, \lambda, T_3+\sigma)} |\operatorname{Rm}| \le \Lambda'$ for all $\lambda > 0$, gives a contradiction. □

Returning to the original sequence $\{\mathcal{M}^\alpha, (p^\alpha, T_0^\alpha)\}_{\alpha=1}^\infty$ and its rescaling $\{\widehat{\mathcal{M}}^\alpha, (\hat{p}^\alpha, 0)\}_{\alpha=1}^\infty$, we can now take a subsequence that converges on the time interval $[0, T_2)$, again necessarily to a standard solution. Then there will be an infinite subsequence $\{\widehat{\mathcal{M}}^{\alpha_\beta}, (\hat{p}^{\alpha_\beta}, 0)\}_{\beta=1}^\infty$ of $\{\widehat{\mathcal{M}}^\alpha, (\hat{p}^\alpha, 0)\}_{\alpha=1}^\infty$, with $\lim_{\beta\to\infty}(T_1^{\alpha_\beta} - T_0^{\alpha_\beta})(\hat{h}^{\alpha_\beta})^{-2} = T_2$, so that $P(\hat{p}^{\alpha_\beta}, 0, A, (T_1^{\alpha_\beta} - T_0^{\alpha_\beta})(\hat{h}^{\alpha_\beta})^{-2})$ is A^{-1}-close to a pointed subset of a standard solution (by the canonical neighborhood assumption, equation (69.2) and Appendix D). This is a contradiction. □

We now finish the proof of Lemma 74.1. If the solution is unscathed on $P(p, T_0, A\hat{h}, T_1 - T_0)$ then we can apply Lemma 74.2 to see that we are in case (1) of the conclusion of Lemma 74.1. Suppose, on the other hand, that the solution is scathed on $P(p, T_0, A\hat{h}, T_1 - T_0)$.

Let t^+ be the largest t so that the solution is unscathed on $P(p, T_0, A\hat{h}, t - T_0)$. We can apply Lemma 74.2 to see that conclusion (1) of Lemma 74.1 holds with T_1 replaced by t^+. As surgery is always performed near the middle of a δ-neck, if $\hat{\delta} << A^{-1}$ then the final time slice in the parabolic neighborhood $P(p, T_0, A\hat{h}, t^+ - T_0)$ cannot intersect a 2-sphere where a surgery is going to be performed. The only other possibility is that the entire ball $B(p, T_0, A\hat{h})$ becomes extinct at time t^+. □

75. II.4.6. Curves that penetrate the surgery region

Let $\mathcal{M}$ be a Ricci flow with (r, δ)-cutoff. The next result, Corollary 75.1, says that if δ is sufficiently small then an admissible curve γ which comes close to a surgery cap at a surgery time will have a large value of $\int_\gamma (R(\gamma(t)) + |\dot{\gamma}(t)|^2)\, dt$. Note that the latter quantity is not quite the same as $\mathcal{L}(\gamma)$, and is invariant under parabolic rescaling.

Corollary 75.1 is used in the extension of Theorem 26.2 to Ricci flows with surgery. The idea is that if δ is small and $L(x,t)$ isn't too large then any $\mathcal{L}$-minimizing sequence of admissible curves joining the basepoint (x_0, t_0) to (x, t) must avoid surgery regions, and will therefore accumulate on a minimizing $\mathcal{L}$-geodesic.

Corollary 75.1. *(cf. Corollary II.4.6) For any $l < \infty$ and $\hat{r} > 0$, we can find $A = A(l, \hat{r}) < \infty$ and $\theta = \theta(l, \hat{r})$ with the following property. Suppose that we are in the situation of Lemma 74.1, with $\delta(T_0) < \hat{\delta}(A, \theta, \hat{r})$. As usual, $\hat{h}$ will be the surgery scale coming from Lemma 71.1. Let $\gamma : [T_0, T_\gamma] \rightarrow \mathcal{M}$ be an admissible curve, with $T_\gamma \in (T_0, T_1]$. Suppose that $\gamma(T_0) \in B(p, T_0, \frac{A\hat{h}}{2})$, $\gamma([T_0, T_\gamma)) \subset P(p, T_0, A\hat{h}, T_\gamma - T_0)$, and either*

a. $T_\gamma = T_1 = T_0 + \theta(\hat{h})^2$,

or

b. $\gamma(T_\gamma) \in \partial B(p, T_0, A\hat{h}) \times [T_0, T_\gamma]$.

Then

$$\int_{T_0}^{T_\gamma} \left(R(\gamma(t), t) + |\dot{\gamma}(t)|^2\right) dt \; > \; l. \tag{75.2}$$

Proof. For the moment, fix $A < \infty$ and $\theta \in (0,1)$. Choose $\hat{\delta} = \hat{\delta}(A, \theta, \hat{r})$ so as to satisfy Lemma 74.1. Let $\mathcal{M}$, (p, T_0), etc., be as in the hypotheses of Lemma 74.1. Let $\gamma : [T_0, T_\gamma] \rightarrow \mathcal{M}$ be a curve as in the hypotheses of the Corollary. From Lemma 74.1, we know that there is a standard solution $\mathcal{S}$ such that the parabolic region $P(p, T_0, A\hat{h}, T_\gamma - T_0) \subset \mathcal{M}$, with basepoint (p, T_0), is (after parabolic rescaling by $\hat{h}^{-2}$) A^{-1}-close to a pointed flow $U_0 \times [0, \hat{T}_\gamma] \subset \mathcal{S}$, the latter having basepoint $(c, 0)$. Here $U_0 \subset \mathcal{S}_0$ and $\hat{T}_\gamma = (T_\gamma - T_0)\hat{h}^{-2}$. Then the image of γ, under the diffeomorphism implicit in the definition of A^{-1}-closeness, gives rise to a smooth curve $\gamma_0 : [0, \hat{T}_\gamma] \rightarrow U_0 \times [0, \hat{T}_\gamma]$ so that (if A is sufficiently large) :

$$\gamma_0(0) \in B(c, 0, \frac{3}{5}A), \tag{75.3}$$

$$\int_{T_0}^{T_\gamma} |\dot{\gamma}|^2 dt \geq \frac{1}{2}\int_0^{\hat{T}_\gamma} |\dot{\gamma}_0|^2 dt, \tag{75.4}$$

$$\int_{T_0}^{T_\gamma} R(\gamma(t), t) dt \geq \frac{1}{2}\int_0^{\hat{T}_\gamma} R(\gamma_0(t), t) dt, \tag{75.5}$$

and

(a) $\hat{T}_\gamma = \theta$,

or

(b) $\gamma_0(\hat{T}_\gamma) \notin P(c, 0, \frac{4}{5}A, \hat{T}_\gamma)$.

In case (a) we have, by Lemma 63.1,

$$\int_{T_0}^{T_\gamma} R(\gamma(t), t) dt \geq \frac{1}{2}\int_0^{\theta} R(\gamma_0(t), t) dt \geq \frac{1}{2}\int_0^{\theta} \text{const.}(1-t)^{-1} dt = \text{const.} \log(1-\theta). \tag{75.6}$$

If we choose θ sufficiently close to 1 then in this case, we can ensure that

$$\int_{T_0}^{T_\gamma} \left(R(\gamma(t), t) + |\dot{\gamma}(t)|^2\right)\, dt \;\geq\; \int_{T_0}^{T_\gamma} R(\gamma(t), t)\, dt \;>\; l. \tag{75.7}$$

In case (b), we may use the fact that the Ricci curvature of the standard solution is everywhere nonnegative, and hence the metric tensor is nonincreasing with time. So if $\pi : \mathcal{S} = \mathcal{S}_0 \times [0,1) \to \mathcal{S}_\theta$ is projection to the time-θ slice and we put $\eta = \pi \circ \gamma_0$ then

$$\begin{aligned} \int_{T_0}^{T_\gamma} |\dot{\gamma}(t)|^2 dt \;&\geq\; \frac{1}{2}\int_0^{\hat{T}_\gamma} |\dot{\gamma_0}(t)|^2 dt \;\geq\; \frac{1}{2}\int_0^{\hat{T}_\gamma} |\dot{\eta}(t)|^2 dt \;\geq\; \frac{1}{2\hat{T}_\gamma}\left(d(\eta(0), \eta(\hat{T}_\gamma))\right)^2 \\ &\geq \frac{1}{2}\left(d(\eta(0), \eta(\hat{T}_\gamma))\right)^2. \end{aligned} \tag{75.8}$$

With our given value of θ, in view of (b), if we take A large enough then we can ensure that $\frac{1}{2}\left(d(\eta(0), \eta(\hat{T}_\gamma))\right)^2 \;>\; l$. This proves the lemma. □

76. II.4.7. A technical estimate

The next result is a technical result that will not be used in the sequel.

Corollary 76.1. *(cf. Corollary II.4.7) For any $Q < \infty$ and $\hat{r} > 0$, there is a $\theta = \theta(Q, \hat{r}) \in (0,1)$ with the following property. Suppose that we are in the situation of Lemma 74.1, with $\delta(T_0) < \hat{\delta}(A, \theta, \hat{r})$ and $A > \epsilon^{-1}$. If $\gamma : [T_0, T_x] \to \mathcal{M}$ is a static curve starting in $B(p, T_0, A\hat{h})$, and*

$$Q^{-1} R(\gamma(t)) \leq R(\gamma(T_x)) \leq Q(T_x - T_0)^{-1} \tag{76.2}$$

for all $t \in [T_0, T_x]$, then $T_x \leq T_0 + \theta\hat{h}^2$.

Remark 76.3. The hypothesis (76.2) in the corollary means that in the scale of the scalar curvature $R(\gamma(T_x))$ at the endpoint $\gamma(T_x)$, the scalar curvature on γ is bounded and the elapsed time of γ is bounded. The conclusion says that given these bounds, the elapsed time is strictly less than that of the corresponding rescaled standard solution.

Proof. If $T_x > T_0 + \theta \hat{h}^2$ then by Lemma 63.1 and 74.1,

$$R(\gamma(T_0 + \theta \hat{h}^2)) \geq \text{const.} (1-\theta)^{-1} \hat{h}^{-2}. \tag{76.4}$$

Thus by (76.2) we get

$$Q^{-1} \, \text{const.} (1-\theta)^{-1} \hat{h}^{-2} \leq R(\gamma(T_x)) \leq Q(T_x - T_0)^{-1}, \tag{76.5}$$

or

$$T_x - T_0 \leq \text{const.} \, Q^2 (1-\theta) \hat{h}^2, \tag{76.6}$$

If we choose θ close enough to 1 then const. $Q^2(1-\theta)\hat{h}^2$ is less than $\theta \hat{h}^2$, which gives a contradiction. □

77. II.5. Statement of the the existence theorem for Ricci flow with surgery

Our presentation of this material follows Perelman's, except for some shuffling of the material. We will be using some terminology introduced in Section 68, as well as results about the L-function and noncollapsing from Sections 78 and 79.

Definition 77.1. A compact Riemannian 3-manifold is *normalized* if $|\operatorname{Rm}| \leq 1$ everywhere, and the volume of every unit ball is at least half the volume of the Euclidean unit ball.

We will use the fact that a smooth normalized Ricci flow, with bounded curvature on compact time intervals, satisfies the Hamilton-Ivey pinching condition of Definition B.5.

The main result of the surgery procedure is Proposition 77.2 (cf. II.5.1), which implies that one can choose positive nonincreasing functions $r : \mathbb{R}_+ \to (0, \infty)$, $\delta : \mathbb{R}_+ \to (0, \infty)$ such that the Ricci flow with (r, δ)-surgery flow starting with any normalized initial condition will be defined for all time.

The actual statement is structured to facilitate a proof by induction:

Proposition 77.2. *(cf. Proposition II.5.1) There exist decreasing sequences $0 < r_j < \epsilon^2$, $\kappa_j > 0$, $0 < \bar{\delta}_j < \epsilon^2$ for $1 \leq j < \infty$, such that for any normalized initial data and any nonincreasing function $\delta : [0, \infty) \to (0, \infty)$ such that $\delta < \bar{\delta}_j$ on $[2^{j-1}\epsilon, 2^j \epsilon]$, the Ricci flow with (r, δ)-cutoff is defined for all time and is κ-noncollapsed at scales below ϵ.*

Here, and in the rest of this section, r and κ will always denote functions defined on an interval $[0, T] \subseteq [0, \infty)$ with the property that $r(t) = r_j$ and $\kappa(t) = \kappa_j$ for all $t \in [0, T] \cap [2^{j-1}\epsilon, 2^j \epsilon)$. By "$\kappa$-noncollapsed at scales below ϵ", we mean that for each $\rho < \epsilon$ and all $(x, t) \in \mathcal{M}$ with $t \geq \rho^2$, whenever $P(x, t, \rho, -\rho^2)$ is unscathed and $|\operatorname{Rm}| \leq \rho^{-2}$ on $P(x, t, \rho, -\rho^2)$, then we also have $\operatorname{vol}(B(x, t, \rho)) \geq \kappa(t) \rho^3$.

Recall that ϵ is a "global" parameter which is assumed to be small, i.e. all statements involving ϵ (explicitly or otherwise) are true provided ϵ is sufficiently small. Proposition 77.2

does not impose any serious new constraints on ϵ. For example, instead of using the time intervals $\{[2^{j-1}\epsilon, 2^j\epsilon]\}_{j=1}^{\infty}$, we could have taken any collection of adjoining time intervals starting at a small positive time. Also, we just need some fixed upper bound on r_j and $\overline{\delta}_j$. We will follow [52] and write these somewhat arbitrary constants in terms of the single global parameter ϵ. Note also that having normalized initial data sets a length scale for the Ricci flow.

The phrase "the Ricci flow with (r, δ)-cutoff is defined for all time" allows for the possibility that the entire manifold goes extinct, i.e. that after some time we are talking about the flow on the empty set.

In the rest of this section we give a sketch of the proof. The details are in the subsequent sections.

Given positive nonincreasing functions r and δ, if one has a normalized initial condition $(M, g(0))$ then there will be a maximal time interval on which the Ricci flow with (r, δ)-cutoff is defined. This interval can be finite only if it is of the form $[0, T)$ for some $T < \infty$, and the Ricci flow with (r, δ)-cutoff on $[0, T)$ extends to a Ricci flow with surgery on $[0, T]$ for which the r-canonical neighborhood assumption fails at time T; see Lemma 73.7. The main point here is that the r-canonical neighborhood assumption allows one to run the flow forward up to the singular time, and then perform surgery, while volume considerations rule out an accumulation of surgery times. Thus the crux of the proof is showing that the functions r and δ can be chosen so that the r-canonical neighborhood assumption will continue to hold, and the Ricci flow with surgery satisfies a noncollapsing condition.

The strategy is to argue by induction on i that r_i, $\bar{\delta}_i$, and κ_i can be chosen (and $\bar{\delta}_{i-1}$ can be adjusted) so that the statement of the proposition holds on the the finite time interval $[0, 2^i\epsilon]$. In the induction step, one establishes the canonical neighborhood assumption using an argument by contradiction similar to the proof of Theorem 52.7. (We recommend that the reader review this before proceeding). The main difference between the proof of Theorem 52.7 and that of Proposition 77.2 is that the non-collapsing assumption, the key ingredient that allows one to implement the blowup argument, is no longer available as a direct consequence of Theorem 26.2, due to the presence of surgeries.

We now discuss the augmentations to the non-collapsing argument of Theorem 26.2 necessitated by surgery; this is treated in detail in sections 78 and 79. We first recall Theorem 26.2 and its proof: if a parabolic ball $P(x_0, t_0, r_0, -r_0^2)$ in Ricci flow (without surgery) is sufficiently collapsed then one uses the L-function with basepoint (x_0, t_0), and the $\mathcal{L}$-exponential map based at (x_0, t_0), to get a contradiction. One considers the reduced volume of a suitably chosen time slice $\mathcal{M}_t$. There is a positive lower bound on the reduced volume coming from the selection of a point where the reduced distance is at most $\frac{3}{2}$, which in turn comes from an application of the maximum principle to the L-function. On the other hand, there is an upper bound on the reduced volume, which the collapsing forces to be small, thereby giving the contradiction. The upper bound comes from the monotonicity of the weighted Jacobian of the $\mathcal{L}$-exponential map. In fact, this upper bound works without significant modification in the presence of surgery, provided one considers only the reduced volume contributed by those points in the time t slice which may be joined to (x_0, t_0) by minimizing $\mathcal{L}$-geodesics lying in the regular part of spacetime (see Lemma 78.11).

To salvage the lower bound on the reduced volume, the basic idea is that by making the surgery parameter δ small, one can force the $\mathcal{L}$-length of any curve passing close to the surgery locus to be large (Lemma 79.3). This implies that if (x,t) is a point where L isn't too large, then there will necessarily be an $\mathcal{L}$-geodesic from (x_0,t_0) to (x,t). To construct the minimizer, one takes a sequence of admissible curves from (x,t) to (x_0,t_0) with $\mathcal{L}$-length tending to the infimum, and argues that they must stay away from the surgeries; hence they remain in a compact part of spacetime, and subconverge to a minimizer. Therefore the calculations from Sections 15-26 will be valid near such a point (x,t). The maximum principle can then be applied as before to show that the minimum of the reduced length is $\le \frac{3}{2}$ on each time slice (see Lemma 78.6).

To be more precise, if one makes the surgery parameter $\delta(t')$ small for a surgery at a given time t' then one can force the $\mathcal{L}$-length of any curve passing close to the time-t' surgery locus to be large, provided that the endtime t_0 of the curve is not too large compared to t'. (If t_0 is much larger than t' then the curve may spend a long time in regions of negative scalar curvature after time t'. The ensuing negative effect on $\mathcal{L}$ could overcome the positive effect of the small surgery parameter.) In the proof of Theorem 26.2, in order to show noncollapsing at time t_0, one went all the way back to a time slice near the initial time and found a point there where l was at most $\frac{3}{2}$. There would be a problem in using this method for Ricci flows with surgery - we would have to constantly redefine $\delta(t')$ to handle the case of larger and larger t_0. The resolution is to not go back to a time slice near the initial time slice. Instead, in order to show κ-noncollapsing in the time slice $[2^i\epsilon, 2^{i+1}\epsilon]$, we will want to get a lower bound on the reduced volume for a time t-slice with t lying in the preceding time interval $[2^{i-1}\epsilon, 2^i\epsilon]$. As we inductively have control over the geometry in the time slice $[2^{i-1}\epsilon, 2^i\epsilon]$, the argument works equally well.

Finally, as mentioned, after obtaining the *a priori* κ-noncollapsing estimate on the interval $[2^i\epsilon, 2^{i+1}\epsilon]$, one proves that the r-canonical neighborhood assumption holds at time $T \in [2^i\epsilon, 2^{i+1}\epsilon]$. One difference here is that because of possible nearby surgeries, there are two ways to obtain the canonical neighborhood : either from closeness to a κ-solution, as in the proof of Theorem 52.7, or from closeness to a standard solution.

78. The L-function of I.7 and Ricci flows with surgery

In this section we examine several points which arise when one adapts the noncollapsing argument of Theorem 26.2 to Ricci flows with surgery. This material is implicit background for Lemma 79.12 and Proposition 84.1. We will use notation and terminology introduced in Section 68.

Let $\mathcal{M}$ be a Ricci flow with surgery, and fix a point $(x_0,t_0) \in \mathcal{M}$. One may define the $\mathcal{L}$-length of an admissible curve γ from (x_0,t_0) to some (x,t), for $t < t_0$, using the formula

$$\mathcal{L}(\gamma) = \int_t^{t_0} \sqrt{t_0 - \bar{t}}\, \left(R + |\dot{\gamma}|^2\right) d\bar{t}, \tag{78.1}$$

where $\dot{\gamma}$ denotes the spatial part of the velocity of γ. One defines the L-function on $\mathcal{M}_{(-\infty,t_0)}$ by setting $L(x,t)$ to be the infimal $\mathcal{L}$-length of the admissible curves from (x_0,t_0) to (x,t) if such an admissible curve exists, and infinity otherwise. We note that if (x,t) is in a surgery

time slice $\mathcal{M}_t^-$ and is actually removed by the surgery then there will not be an admissible curve from (x_0, t_0) to (x, t).

If γ is an admissible curve lying in $\mathcal{M}_{\mathrm{reg}}$ then the first variation formula applies. Hence an admissible curve in $\mathcal{M}_{\mathrm{reg}}$ from (x_0, t_0) to (x, t) whose $\mathcal{L}$-length equals $L(x, t)$ will satisfy the $\mathcal{L}$-geodesic equation. If γ is a stable $\mathcal{L}$-geodesic in $\mathcal{M}_{\mathrm{reg}}$ then the proof of the monotonicity along γ of the weighted Jacobian $\tau^{-\frac{3}{2}} \exp(-l(\tau)) J(\tau)$ remains valid. Similarly, if $U \subset \mathcal{M}_{(-\infty,t_0)}$ is an open set such that every $(x, t) \in U$ is accessible from (x_0, t_0) by a minimizing $\mathcal{L}$-geodesic (i.e. an $\mathcal{L}$-geodesic of $\mathcal{L}$-length $L(x, t)$) contained in $\mathcal{M}_{\mathrm{reg}}$, then the arguments of Section 24 imply that the differential inequality

$$\bar{L}_\tau + \Delta \bar{L} \leq 6 \tag{78.2}$$

holds in U, in the barrier sense, where $\tau = t_0 - t$, $\bar{L} = 2\sqrt{\tau}\, L$ and $l = \frac{\bar{L}}{4\tau}$.

Lemma 78.3 (Existence of $\mathcal{L}$-minimizers)**.** *Let $\mathcal{M}$ be a Ricci flow with surgery defined on $[a, b]$. Suppose that $(x_0, t_0) \in \mathcal{M}$ lies in the backward time slice $\mathcal{M}_{t_0}^-$.*

(1) For each $(x, t) \in \mathcal{M}_{[a,t_0)}$ with $L(x, t) < \infty$, there exists an $\mathcal{L}$-minimizing admissible path $\gamma : [t, t_0] \to \mathcal{M}$ from (x, t) to (x_0, t_0) which satisfies the $\mathcal{L}$-geodesic equation at every time $\bar{t} \in (t, t_0)$ for which $\gamma(\bar{t}) \in \mathcal{M}_{\mathrm{reg}}$.

(2) L is lower semicontinuous on $\mathcal{M}_{[a,t_0)}$ and continuous on $\mathcal{M}_{\mathrm{reg}} \cap \mathcal{M}_{[a,t_0)}$. (Note that $\mathcal{M}_a^+ \subset \mathcal{M}_{\mathrm{reg}}$.)

(3) Every sequence $(x_j, t_j) \in \mathcal{M}_{[a,t_0)}$ with $\limsup_j L(x_j, t_j) < \infty$ *has a convergent subsequence.*

Proof. (1) Let $\{\gamma_j : [t, t_0] \to \mathcal{M}\}_{j=1}^\infty$ be a sequence of admissible curves from (x, t) to (x_0, t_0) such that $\lim_{j\to\infty} \mathcal{L}(\gamma_j) = L(x, t) < \infty$. By restricting the sequence, we may assume that $\sup_j \mathcal{L}(\gamma_j) < 2L(x, t)$. We claim that there is a subsequence of the γ_j's that

(a) converges uniformly to some $\gamma_\infty : [t, t_0] \to \mathcal{M}$,

and

(b) converges weakly to γ_∞ in $W^{1,2}$ on any subinterval $[t', t''] \subset [t, t_0)$ such that $[t', t'']$ is free of singular times.

To see this, note that on any time interval $[c, d] \subset [t, t_0)$ which is free of singular times, one may apply the Schwarz inequality to the $\mathcal{L}$-length, along with the fact that the metrics on the time slices $\mathcal{M}_t$, $t \in [c, d]$, are uniformly biLipschitz to each other, to conclude that the γ_j's are uniformly Hölder-continuous on $[c, d]$. We know that $\gamma_j(t')$ lies in $\mathcal{M}_{t'}^- \cap \mathcal{M}_{t'}^+$ for each surgery time $t' \in (t, t_0)$, and so one can use similar reasoning to get Hölder control on a short time interval of the form $[t'', t']$. Using a change of variable as in (17.6), one obtains uniform Hölder control near t_0 after reparametrizing with s. It follows that the γ_j's are equicontinuous and map into a compact part of spacetime, so Arzela-Ascoli applies; therefore, by passing to a subsequence we may assume that (a) holds.

To show (b), we apply weak compactness to the sequence

$$\{\gamma_j\big|_{[t',t'']}\}; \tag{78.4}$$

this is justified by the fact that the paths $\gamma_j\Big|_{[t',t'']}$'s remain in a part of $\mathcal{M}$ with bounded geometry. Thus we may assume that our sequence $\{\gamma_j\}$ converges uniformly on $[t,t_0]$ and weakly on every subinterval $[t',t'']$ as in (b). By weak lower semicontinuity of $\mathcal{L}$-length, it follows that the $W^{1,2}$-path γ_∞ has $\mathcal{L}$-length $\leq L(x,t)$. Since any $W^{1,2}$ path may be approximated in $W^{1,2}$ by admissible curves with the same endpoints, it follows that γ_∞ minimizes $\mathcal{L}$-length among $W^{1,2}$ paths, and therefore it restricts to a smooth solution of the $\mathcal{L}$-geodesic equation on each time interval $[t',t''] \subset [t,t_0]$ such that (t',t'') is free of singular times. Hence γ_∞ is an $\mathcal{L}$-minimizing admissible curve.

(2) Pick $(x,t) \in \mathcal{M}_{[a,t_0)}$. To verify lower semicontinuity at (x,t) we suppose the sequence $\{(x_j,t_j)\} \subset \mathcal{M}_{[a,t_0)}$ converges to (x,t) and $\liminf_{j\to\infty} L(x_j,t_j) < \infty$. By (1) there is a sequence $\{\gamma_j\}$ of $\mathcal{L}$-minimizing admissible curves, where γ_j runs from (x_j,t_j) to (x_0,t_0). By the reasoning above, a subsequence of $\{\gamma_j\}$ converges uniformly and weakly in $W^{1,2}$ to a $W^{1,2}$ curve $\gamma_\infty : [t,t_0] \to \mathcal{M}$ going from (x,t) to (x_0,t_0), with

$$\mathcal{L}(\gamma_\infty) \leq \liminf_{j\to\infty} \mathcal{L}(\gamma_j). \tag{78.5}$$

Therefore $L(x,t) \leq \liminf_{j\to\infty} L(x_j,t_j)$, and we have established semicontinuity. If $(x,t) \in \mathcal{M}_{\mathrm{reg}}$, the opposite inequality obviously holds, so in this case (x,t) is a point of continuity.

(3) Because $\{L(x_j,t_j)\}$ is uniformly bounded, any sequence $\{\gamma_j\}$ of $\mathcal{L}$-minimizing paths with $\gamma_j(t_j) = (x_j,t_j)$ will be equicontinuous, and hence by Arzela-Ascoli a subsequence converges uniformly. Therefore a subsequence of $\{(x_j,t_j)\}$ converges. $\square$

The fact that (78.2) can hold locally allows one to appeal – under appropriate conditions – to the maximum principle as in Section 24 to prove that $\min l \leq \frac{3}{2}$ on every time slice. Recall that $l = \frac{L}{2\sqrt{\tau}} = \frac{\overline{L}}{4\tau}$.

Lemma 78.6. *Suppose that $\mathcal{M}$ is a Ricci flow with surgery defined on $[a,b]$. Take $t_0 \in (a,b]$ and $(x_0,t_0) \in \mathcal{M}^-_{t_0}$. Suppose that for every $t \in [a,t_0)$, every admissible curve $[t,t_0] \to \mathcal{M}$ ending at (x_0,t_0) which does not lie in $\mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}^-_{t_0}$ has reduced length strictly greater than $\frac{3}{2}$. Then there is a point $(x,a) \in \mathcal{M}^+_a$ where $l(x,a) \leq \frac{3}{2}$.*

Remark 78.7. In the lemma we consider the Ricci flow with surgery to begin at time a. Hence $\mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}^-_{t_0} = \mathcal{M}^+_a \cup \mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}^-_{t_0}$ and so the hypothesis of the lemma is a statement about the reduced lengths of barely admissible curves, in the sense of Section 68.

Proof. As in the case when there are no surgeries, the proof relies on the maximum principle and a continuity argument.

Let $\beta : \mathcal{M}_{[a,t_0)} \to \mathbb{R} \cup \{\infty\}$ be the function

$$\beta = \bar{L} - 6\tau = 4\tau\left(l - \frac{3}{2}\right), \tag{78.8}$$

where as usual, $\tau(x,t) = t_0 - t$. Note that for each $\tau \in (0, t_0 - a]$, the function β attains a minimum $\beta_{\min}(\tau) < \infty$ on the slice $\mathcal{M}_{t_0-\tau}$, because by (2) of Lemma 78.3, it is continuous on the compact manifold $\mathcal{M}^+_{t_0-\tau}$ (as seen by changing the parameter a of Lemma 78.3 to $t_0 - \tau$), and $\beta \equiv \infty$ on $\mathcal{M}_{t_0-\tau} - \mathcal{M}^+_{t_0-\tau}$. Thus it suffices to show that $\beta_{\min}(t_0 - a) \leq 0$.

From Lemma 24.3, $\beta_{\min}(\tau) < 0$ for $\tau > 0$ small. Let $\tau_1 \in (0, t_0 - a]$ be the supremum of the $\bar{\tau} \in (0, t_0 - a]$ such that $\beta_{\min} < 0$ on the interval $(0, \bar{\tau})$.

We claim that

(a) $\beta_{\min}$ is continuous on $(0, \tau_1)$

and

(b) The upper right τ-derivative of $\beta_{\min}$ is nonpositive on $(0, \tau_1)$.

To see (a), pick $\tau \in (0, \tau_1)$, suppose that $\{\tau_j\} \subset (0, \tau_1)$ is a sequence converging to τ and choose $(x_j, t_0 - \tau_j) \in \mathcal{M}^+_{t_0 - \tau_j}$ such that $\beta(x_j, t_0 - \tau_j) = \beta_{\min}(\tau_j) < 0$. By Lemma 78.3 part (3), the sequence $\{(x_j, t_0 - \tau_j)\}$ subconverges to some $(x, t_0 - \tau) \in \mathcal{M}^+_{t_0 - \tau}$ for which $\beta(x, t_0 - \tau) \le \liminf_{j \to \infty} \beta(x_j, t_0 - \tau_j)$. Thus $\beta_{\min}$ is lower semicontinuous at τ. On the other hand, since $\beta_{\min}(\tau) < 0$, the minimum of β on $\mathcal{M}_{t_0 - \tau}$ will be attained at a point $(x, t_0 - \tau) \in \mathcal{M}^+_{t_0 - \tau}$ lying in the interior of $\mathcal{M}^-_{t_0 - \tau} \cap \mathcal{M}^+_{t_0 - \tau}$, as $\beta > 0$ elsewhere on $\mathcal{M}_{t_0 - \tau}$ (by Lemma 78.3 and the hypothesis on admissible curves). Therefore β is continuous at $(x, t_0 - \tau)$, which implies that $\beta_{\min}$ is upper semicontinuous at τ. This gives (a).

Part (b) of the claim follows from the fact that if $\tau \in (0, \tau_1)$ and the minimum of β on $\mathcal{M}_{t_0 - \tau}$ is attained at $(x, t_0 - \tau)$ then $l(x, \tau) < \frac{3}{2}$, so there is a neighborhood U of $(x, t_0 - \tau)$ such that the inequality

$$\frac{\partial \beta}{\partial \tau} + \Delta \beta \le 0 \tag{78.9}$$

holds in the barrier sense on U (by Lemma 78.3 and the hypothesis on admissible curves). Hence the upper right derivative $\frac{d}{ds}\Big|_{s=0} \beta(x, \tau + s)$ is nonpositive, so the upper right τ-derivative of $\beta_{\min}(\tau)$ is also nonpositive.

The claim implies that $\beta_{\min}$ is nonincreasing on $(0, \tau_1)$, and so $\limsup_{\tau \to \tau_1^-} \beta(\tau) < 0$. By parts (2) and (3) of Lemma 78.3, we have $\beta_{\min}(\tau_1) < 0$, and the minimum is attained at some $(x, t_0 - \tau_1) \in \mathcal{M}_{\mathrm{reg}}$. (Recall that $\mathcal{M}_a \subset \mathcal{M}_{\mathrm{reg}}$.) This implies that $\tau_1 = t_0 - a$, for otherwise $\beta_{\min}(\tau)$ would be strictly negative for $\tau \ge \tau_1$ close to τ_1, contradicting the definition of τ_1. $\square$

The notion of local collapsing can be adapted to Ricci flows with surgery, as follows.

Definition 78.10. Let $\mathcal{M}$ be a Ricci flow with surgery defined on $[a, b]$. Suppose that $(x_0, t_0) \in \mathcal{M}$ and $r > 0$ are such that $t_0 - r^2 \ge a$, $B(x_0, t_0, r) \subset \mathcal{M}^-_{t_0}$ is a proper ball and the parabolic ball $P(x_0, t_0, r, -r^2)$ is unscathed. Then $\mathcal{M}$ *is κ-collapsed at (x_0, t_0) at scale r* if $|\operatorname{Rm}| \le r^{-2}$ on $P(x_0, t_0, r, -r^2)$ and $\operatorname{vol}(B(x_0, t_0, r)) < \kappa r^3$; otherwise it is κ-noncollapsed.

We make use of the following variant of the noncollapsing argument from Section 26.

Lemma 78.11. *(Local version of reduced volume comparison) There is a function $\kappa' : \mathbb{R}_+ \to \mathbb{R}_+$, satisfying $\lim_{\kappa \to 0} \kappa'(\kappa) = 0$, with the following property. Let $\mathcal{M}$ be a Ricci flow with surgery defined on $[a, b]$. Suppose that we are given $t_0 \in (a, b]$, $(x_0, t_0) \in \mathcal{M}_{t_0} \cap \mathcal{M}_{\mathrm{reg}}$, $t \in [a, t_0)$ and $r \in (0, \sqrt{t_0 - t})$. Let Y be the set of points $(x, t) \in \mathcal{M}_t$ that are accessible from (x_0, t_0) by means of minimizing $\mathcal{L}$-geodesics which remain in $\mathcal{M}_{\mathrm{reg}}$. Assume in addition that $\mathcal{M}$ is κ-collapsed at (x_0, t_0) at scale r, i.e. $P(x_0, t_0, r, -r^2) \cap \mathcal{M}_{[t_0 - r^2, t_0)} \subset \mathcal{M}_{\mathrm{reg}}$, $|\operatorname{Rm}| \le r^{-2}$*

on $P(x_0, t_0, r, -r^2)$, and $\mathrm{vol}(B(x_0, t_0, r)) < \kappa r^3$. *Then the reduced volume of Y is at most $\kappa'(\kappa)$.*

Proof. Let $\hat{Y} \subset T_{x_0}\mathcal{M}_{t_0}$ be the set of vectors $v \in T_{x_0}\mathcal{M}_{t_0}$ such that there is a minimizing $\mathcal{L}$-geodesic $\gamma : [t, t_0] \to \mathcal{M}_{\mathrm{reg}}$ running from (x_0, t_0) to some point in Y, with

$$\lim_{\bar{t} \to t_0} \sqrt{t_0 - \bar{t}}\, \dot{\gamma}(\bar{t}) = -v. \tag{78.12}$$

The calculations from Sections 17-23 apply to $\mathcal{L}$-geodesics sitting in $\mathcal{M}_{\mathrm{reg}}$. In particular, the monotonicity of the weighted Jacobian $\tau^{-\frac{n}{2}} \exp(-l(\tau))J(\tau)$ holds. Now one repeats the proof of Theorem 26.2, working with the set $\hat{Y}$ instead of the set of initial velocities of *all* minimizing $\mathcal{L}$-geodesics. □

79. Establishing noncollapsing in the presence of surgery

The key result of this section, Lemma 79.12, gives conditions under which one can deduce noncollapsing on a time interval I_2, given a noncollapsing bound on a preceding interval I_1 and lower bounds on r on $I_1 \cup I_2$.

Definition 79.1. The $\mathcal{L}_+$*-length* of an admissible curve γ is

$$\mathcal{L}_+(\gamma, \tau) \;=\; \int_{t_0-\tau}^{t_0} \sqrt{t_0 - t}\; \left(R_+(\gamma(t), t) \;+\; |\dot{\gamma}(t)|^2 \right)\, dt, \tag{79.2}$$

where $R_+(x, t) = \max(R(x, t), 0)$.

Lemma 79.3. *(Forcing $\mathcal{L}_+$ to be large, cf. Lemma II.5.3)*

For all $\Lambda < \infty$, $\bar{r} > 0$ and $\hat{r} > 0$, there is a constant $F_0 = F_0(\Lambda, \bar{r}, \hat{r})$ with the following property. Suppose that

- *$\mathcal{M}$ is a Ricci flow with (r, δ)-cutoff defined on an interval containing $[t, t_0]$, where $r([t, t_0]) \subset [\hat{r}, \epsilon]$,*
- *$r_0 \geq \bar{r}$, $B(x_0, t_0, r_0)$ is a proper ball which is unscathed on $[t_0 - r_0^2, t_0]$, and $|\operatorname{Rm}| \leq r_0^{-2}$ on $P(x_0, t_0, r_0, -r_0^2)$,*
- *$\gamma : [t, t_0] \to \mathcal{M}$ is an admissible curve ending at (x_0, t_0) whose image is not contained in $\mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}_{t_0}$, and*
- *$\delta < F_0(\Lambda, \bar{r}, \hat{r})$ on $[t, t_0]$.*

Then $\mathcal{L}_+(\gamma) > \Lambda$.

Proof. The idea is that the hypotheses on γ imply that it must touch the part of the manifold added during surgery at some time $\bar{t} \in [t, t_0]$. Then either γ has to move very fast at times close to $\bar{t}$ or t_0, or it will stay in the surgery region while it develops large scalar curvature. In the first case $\mathcal{L}_+(\gamma)$ will be large because of the $|\dot{\gamma}|^2$ term in the formula for $\mathcal{L}_+$, and in the second case it will be large because of the $R(\gamma)$ term.

First, we can assume that F_0 is small enough so that $F_0 < \sqrt{\frac{\bar{r}}{100\epsilon}}$. Then since

$$\max_{[t,t_0]} h(t) \le \left(\max_{[t,t_0]} \delta\right)^2 \left(\max_{[t,t_0]} r(t)\right) \le F_0^2 \epsilon, \tag{79.4}$$

we have $\max_{[t,t_0]} h(t) < \frac{\bar{r}}{100} \le \frac{r_0}{100}$.

Put $\Delta t = 10^{-10}\bar{r}^4\Lambda^{-2}$. It always suffices to prove the lemma for a larger value of Λ, so without loss of generality we can assume that $\Delta t \le \bar{r}^2 \le r_0^2$. Set

$$A = A((\Delta t)^{-\frac{1}{2}}\Lambda, \hat{r}), \quad \theta = \theta((\Delta t)^{-\frac{1}{2}}\Lambda, \hat{r}) \tag{79.5}$$

where $A(\cdot,\cdot)$ and $\theta(\cdot,\cdot)$ are the functions from Corollary 75.1. That is, we will eventually be applying Corollary 75.1 with $l = (\Delta t)^{-\frac{1}{2}}\Lambda$. We impose the additional constraint on F_0 that

$$F_0 \le \hat{\delta}(A+2, \theta, \hat{r}) \tag{79.6}$$

on the interval $[t, t_0]$, where $\hat{\delta}$ is the function from Lemma 74.1.

As γ is admissible but is not contained in $\mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}_{t_0}$, it must pass through the boundary of a surgery cap at some time in the interval $[t, t_0)$ or it must start in the interior of a surgery cap at time t. By dropping an initial segment of γ if necessary, we may assume that $\gamma(t)$ lies in a surgery cap.

Let x denote the tip of the surgery cap. Note that

$$P(x_0, t_0, r_0, -r_0^2) \cap P(x, t, Ah(t), \theta h^2(t)) = \emptyset \tag{79.7}$$

since by Lemma 74.1 the scalar curvature on $P(x, t, Ah(t), \theta h^2(t))$ is at least $\frac{h^{-2}}{2} > \frac{10^4}{2} r_0^{-2}$, while $|\operatorname{Rm}| \le r_0^{-2}$ on $P(x_0, t_0, r_0, -r_0^2)$. Therefore when going backward in time from (x_0, t_0), γ must leave the parabolic region $P(x_0, t_0, r_0, -r_0^2)$ before it arrives at (x, t). If it exits at a time $\tilde{t} > t_0 - \Delta t$ then applying the Schwarz inequality we get

$$\int_{\tilde{t}}^{t_0} \sqrt{t_0 - s}\, |\dot{\gamma}(s)|^2 \, ds \ \ge \ \left(\int_{\tilde{t}}^{t_0} |\dot{\gamma}(s)| \, ds\right)^2 \left(\int_{\tilde{t}}^{t_0} (t_0 - s)^{-1/2} \, ds\right)^{-1} \ge \ \frac{1}{100}\, r_0^2 (\Delta t)^{-1/2} > \Lambda, \tag{79.8}$$

where the factor of $\frac{1}{100}$ comes from the length distortion estimate of Section 27, using the fact that $|\operatorname{Rm}| \le r_0^{-2}$ on $P(x_0, t_0, r_0, -r_0^2)$. So we can restrict to the case when γ exits $P(x_0, t_0, r_0, -\Delta t)$ through the initial time slice at time $t_0 - \Delta t$. In particular, by (79.7), γ must exit the parabolic region $P(x, t, Ah(t), \theta h^2(t))$ by time $t_0 - \Delta t$.

By Lemma 74.1, the parabolic region $P(x, t, Ah, \theta h^2)$ is either unscathed, or it coincides (as a set) with the parabolic region $P(x, t, Ah, s)$ for some $s \in (0, \theta h^2)$ and the entire final time slice $P(x, t, Ah, s) \cap \mathcal{M}_{t+s}$ of $P(x, t, Ah, s)$ is thrown away by a surgery at time $t + s$.

One possibility is that γ exits $P(x, t, Ah, \theta h^2)$ through the final time slice. If this is the case then $P(x, t, Ah, \theta h^2)$ must be unscathed (as otherwise the final face is removed by surgery at time $t + s < t + \theta h^2$ and γ would have nowhere to go after this time), so γ lies in $P(x, t, Ah, \theta h^2)$ for the entire time interval $[t, t + \theta h^2]$.

The other possibility is that γ leaves $P(x, t, Ah, \theta h^2)$ before the final time slice of $P(x, t, Ah, \theta h^2)$, in which case it exits the ball $B(x, t, Ah)$ by time $t + \theta h^2$.

Corollary 75.1 applies to either of these two possibilities. Putting

$$T_\gamma = \sup\{\bar{t} \in [t, t+\theta h^2] \mid \gamma([t,\bar{t}]) \subset P(x,t,Ah,\theta h^2)\} \tag{79.9}$$

and using the fact that $T_\gamma \le t_0 - \Delta t$, we have

(79.10)

$$\int_t^{t_0} \sqrt{t_0 - s}\left(R_+(\gamma(s),s) \;+\; |\dot{\gamma}(s)|^2\right)\, ds \;\ge\; \int_t^{T_\gamma} \sqrt{t_0 - s}\left(R_+(\gamma(s),s) \;+\; |\dot{\gamma}(s)|^2\right)\, ds \;\ge$$

$$(\Delta t)^{1/2} \int_t^{T_\gamma} \left(R_+(\gamma(s),s) \;+\; |\dot{\gamma}(s)|^2\right)\, ds \;\ge\; (\Delta t)^{1/2}\, l \;=\; \Lambda,$$

where the last inequality comes from Corollary 75.1 and the choice of A, θ, and δ in (79.5) and (79.6). This completes the proof. □

Lemma 79.11. *If $\mathcal{M}$ is a Ricci flow with surgery, with normalized initial condition at time zero, then for all $t \ge 0$, $R(x,t) \;\ge\; -\frac{3}{2}\,\frac{1}{t+\frac{1}{4}}$.*

Proof. From the initial conditions, $R_{min}(0) \;\ge\; -6$. If the Ricci flow is smooth then (B.2) implies that $R_{min}(t) \;\ge\; -\frac{3}{2}\,\frac{1}{t+\frac{1}{4}}$. If there is a surgery at time t_0 then R_{min} on $\mathcal{M}^+_{t_0}$ equals R_{min} on $\mathcal{M}^-_{t_0}$, as surgery is done in regions of high scalar curvature. The lemma follows by applying (B.2) on the time intervals between the singular times. □

In the statement of the next lemma, one has successive time intervals $[a,b)$ and $[b,c)$. As a mnemonic we use the subscript $-$ for quantities attached to the earlier interval $[a,b)$, and $+$ for those associated with $[b,c)$. We will also assume that the global parameter ϵ is small enough that the Φ-pinching condition implies that whenever $|\operatorname{Rm}(x,t)| \ge \epsilon^{-2}$, then $R(x,t) > \frac{|\operatorname{Rm}(x,t)|}{100}$. (We remind the reader of the role of the parameter ϵ; see Remark 58.5.)

Lemma 79.12. *(Noncollapsing estimate) (cf. Lemma II.5.2)*

Suppose $\epsilon \ge r_- \ge r_+ > 0$, $\kappa_- > 0$, $E_- > 0$ and $E < \infty$. Then there are constants $\overline{\delta} = \overline{\delta}(r_-, r_+, \kappa_-, E_-, E)$ and $\kappa_+ = \kappa_+(r_-, \kappa_-, E_-, E)$ with the following property. Suppose that

- *$a < b < c$, $\quad b - a \ge E_-$, $\quad c - a \le E$,*
- *$\mathcal{M}$ is a Ricci flow with (r,δ)-cutoff with normalized initial condition defined on a time interval containing $[a,c)$,*
- *$r \ge r_-$ on $[a,b)$ and $r \ge r_+$ on $[b,c)$,*
- *$r \le \epsilon$,*
- *$\mathcal{M}$ is κ_--noncollapsed at scales below ϵ on $[a,b)$ and*
- *$\delta \le \overline{\delta}$ on $[a,c)$,*

Then $\mathcal{M}$ is κ_+-noncollapsed at scales below ϵ on $[b,c)$.

Remark 79.13. The important point to notice here is that $\bar{\delta}$ is allowed to depend on the lower bound r_+ on $[b,c)$, but the noncollapsing constant κ_+ does not depend on r_+.

Proof. In the proof, we can assume that $\frac{r_+}{100} \leq \sqrt{E_-/3}$. If this were not the case then we could prove the lemma with r_+ replaced by $100\sqrt{E_-/3}$. Then the lemma would also hold for the original value of r_+.

First, from Lemma 79.11, $R \geq -6$ on $\mathcal{M}_{[a,c)}$.

Suppose that $r_0 \in (0, \epsilon)$, $(x_0, t_0) \in \mathcal{M}_{[b,c)}$, $B(x_0, t_0, r_0)$ is a proper ball unscathed on the interval $[t_0 - r_0^2, t_0]$, and $|\operatorname{Rm}| \leq r_0^{-2}$ on $P(x_0, t_0, r_0, -r_0^2)$.

We first assume that $r_0 \leq \sqrt{E_-/3}$ and $r_0 \geq \frac{r_+}{100}$.

We will consider $\mathcal{L}$-length, $\mathcal{L}_+$-length, etc in $\mathcal{M}_{[a,t_0)}$ with basepoint at (x_0, t_0). Suppose that $\widehat{t} \in [a, t_0)$. Then for any admissible curve $\gamma : [\widehat{t}, t_0] \to \mathcal{M}_{[a,t_0]}$ ending at (x_0, t_0), we have

$$\mathcal{L}(\gamma) \leq \mathcal{L}_+(\gamma) \leq \mathcal{L}(\gamma) + \int_a^c 6\sqrt{c-t}\, dt \leq \mathcal{L}(\gamma) + 4E^{\frac{3}{2}} \tag{79.14}$$

$$\text{and} \quad l(x,t) \geq \frac{L_+ - 4E^{\frac{3}{2}}}{2E^{\frac{1}{2}}}.$$

Assume that $\bar{\delta} \leq F_0(4E^{\frac{1}{2}} + 4E^{\frac{3}{2}}, \frac{r_+}{100}, r_+)$ where F_0 is the function from Lemma 79.3. Then by (79.14) and Lemma 79.3, we conclude that any admissible curve $[\widehat{t}, t_0] \to \mathcal{M}_{[a,t_0]}$ ending at (x_0, t_0) which does not lie in $\mathcal{M}_{\mathrm{reg}} \cup \mathcal{M}_{t_0}$ has reduced length bounded below by $2 = \frac{3}{2} + \frac{1}{2}$. By Lemma 78.6 there is an admissible curve $\gamma : [a, t_0] \to \mathcal{M}$ ending at (x_0, t_0) such that

$$\mathcal{L}(\gamma) = L(\gamma(a)) = 2\sqrt{t_0 - a}\, l(\gamma(a)) \leq 3\sqrt{t_0 - a}, \tag{79.15}$$

so by (79.14) it follows that

$$\mathcal{L}_+(\gamma) \leq 3\sqrt{t_0 - a} + 4E^{\frac{3}{2}} \leq 3\sqrt{E} + 4E^{\frac{3}{2}}. \tag{79.16}$$

Set

$$t_1 = a + \frac{b-a}{3}, \quad t_2 = a + \frac{2(b-a)}{3} \tag{79.17}$$

and

$$\rho = \left(3\sqrt{E} + 4E^{\frac{3}{2}}\right)\left(\frac{1}{3}E_-\right)^{-\frac{3}{2}}. \tag{79.18}$$

By construction, $t_2 \leq t_0 - r_0^2$. Note that there is a $\bar{t} \in [t_1, t_2]$ such that $R(\gamma(\bar{t})) \leq \rho$. Otherwise we would get
(79.19)

$$\mathcal{L}_+(\gamma) > \int_{t_1}^{t_2} \sqrt{t_0 - t}\, R_+(\gamma(t))\, dt \geq \sqrt{\frac{1}{3}E_-} \int_{t_1}^{t_2} \rho\, dt \geq \sqrt{\frac{1}{3}E_-}\left(\frac{1}{3}E_-\right)\rho = 3\sqrt{E} + 4E^{\frac{3}{2}},$$

contradicting (79.16).

Put $\overline{x} = \gamma(\bar{t})$. By Lemma 70.1, there is an estimate of the form

$$R \leq \text{const.}\, s^{-2} \tag{79.20}$$

on the parabolic ball $\hat{P} = P(\overline{x}, \overline{t}, s, -s^2)$ with $s^{-2} = \text{const.}(\rho + r_-^{-2})$. Appealing to Hamilton-Ivey curvature pinching as usual, we get that $|\operatorname{Rm}| \leq \text{const.}\, s^{-2}$ in $\hat{P}$. If $\overline{t} - s^2 < a$ then we shrink s (as little as possible) to ensure that $\hat{P} \subset \mathcal{M}_{[a,b)}$. Provided that $\bar{\delta}$ is less than a small constant $c_1 = c_1(r_-, E_-, E)$, we can guarantee that $\hat{P}$ is unscathed, by forcing the curvature in a surgery cap to exceed our bound (79.20) on R. Put $U = \mathcal{M}_{\bar{t} - \frac{1}{2}s^2} \cap \hat{P}$. If $s < \epsilon$ then the κ_--noncollapsing assumption on $[a, b)$ gives a lower bound on $\operatorname{vol}(B(\bar{x}, \bar{t}, s))s^{-3}$. If $s \geq \epsilon$ then the κ_--noncollapsing assumption gives a lower bound on $\operatorname{vol}(B(\bar{x}, \bar{t}, \epsilon/2))(\epsilon/2)^{-3}$. In either, case, we get a lower bound on $\operatorname{vol}(B(\bar{x}, \bar{t}, s))$ and hence a lower bound $\operatorname{vol}(U) \geq v = v(r_-, \kappa_-, E_-, E)$. Now every point in U can be joined to (x_0, t_0) by a curve of $\mathcal{L}_+$-length at most $\Lambda_+ = \Lambda_+(r_-, E_-, E)$, by concatenating an admissible curve $[\bar{t} - \frac{1}{2}s^2, \bar{t}] \to \mathcal{M}$ (of controlled $\mathcal{L}_+$-length) with $\gamma\Big|_{[\bar{t}, t_0]}$. Shrinking $\bar{\delta}$ again, we can apply Lemmas 78.3 and 79.3 with (79.14) to ensure that every point in U can be joined to (x_0, t_0) by a minimizing $\mathcal{L}$-geodesic lying in $\mathcal{M}_{\text{reg}} \cup \mathcal{M}_{t_0}$. Lemma 78.11 then implies that

$$\operatorname{vol}(B(x_0, t_0, r_0))r_0^{-3} \geq \kappa_1 = \kappa_1(r_-, \kappa_-, E_-, E). \tag{79.21}$$

(We briefly recall the argument. We have a parabolic ball around $(\bar{x}, \bar{t})$, of small but controlled size, on which we have uniform curvature bounds. The lower volume bound coming from the κ_--noncollapsing assumption on $[a, b)$ means that we have bounded geometry on the parabolic ball. As we have a fixed upper bound on $l(\bar{x}, \bar{t})$, we can estimate from below the reduced volume of the accessible points $Y \subset \mathcal{M}^+_{\bar{t} - \frac{1}{2}s^2}$. Then we obtain a lower bound on $\operatorname{vol}(B(x_0, t_0, r_0))r_0^{-3}$ as in Theorem 26.2.)

This completes the proof of the lemma when $r_0 \leq \sqrt{E_-/3}$ and $r_0 \geq \frac{r_+}{100}$.

Now suppose that $r_0 > \sqrt{E_-/3}$. Applying our noncollapsing estimate (79.21) to the ball of radius $\sqrt{E_-/3}$ gives

$$\operatorname{vol}(B(x_0, t_0, r_0))r_0^{-3} \geq \left(\operatorname{vol}(B(x_0, t_0, \sqrt{E_-/3})(E_-/3)^{-\frac{3}{2}}\right) \frac{(E_-/3)^{\frac{3}{2}}}{r_0^3} \geq \kappa_1 \frac{(E_-/3)^{\frac{3}{2}}}{\epsilon^3} = \kappa_2, \tag{79.22}$$

where $\kappa_2 = \kappa_2(r_-, \kappa_-, E_-, E)$.

The next sublemma deals with the case when $r_0 < \frac{r_+}{100}$.

Sublemma 79.23. *If* $r_0 < \frac{r_+}{100}$ *then* $\operatorname{vol}(B(x_0, t_0, r_0))r_0^{-3} \geq \kappa_3 = \kappa_3(r_-, \kappa_-, E_-, E)$.

Proof. Let s be the maximum of the numbers $\bar{s} \in [r_0, \frac{r_+}{100}]$ such that $B(x_0, t_0, \bar{s})$ is unscathed on $[t_0 - \bar{s}^2, t_0]$, and $|\operatorname{Rm}| \leq \bar{s}^{-2}$ on $P(x_0, t_0, \bar{s}, -\bar{s}^{-2})$. Then either

(a) Some point (x, t) on the frontier of $P(x_0, t_0, s, -s^2)$ lies in a surgery cap

or

(b) Some point (x, t) in the closure of $P(x_0, t_0, s, -s^2)$ has $|\operatorname{Rm}| = s^{-2}$

or

(c) $s = \frac{r_+}{100}$.

In case (a) the scalar curvature at (x,t) will satisfy $R(x,t) \in (\frac{h^{-2}}{2}, 10h^{-2})$, since (x,t) lies in the cap at the surgery time. Since $|\operatorname{Rm}(x,t)| \leq s^{-2}$ we conclude that $s \leq \text{const.}\, h(t)$. If $\bar{\delta}$ is small then the pointed time slice $(\mathcal{M}_t, (x,t))$ will be close, modulo scaling by $h(t)$, to the initial condition of the standard solution with the basepoint somewhere in the cap. Using the fact that the time slices of $P(x_0, t_0, s, -s^2)$ have comparable metrics, along with the fact that $r_0 \leq s \leq \text{const.}\, h(t)$, we get a lower bound $\operatorname{vol}(B(x_0,t_0,r_0))r_0^{-3} \geq \text{const.}$

In case (b), we have a static curve $\gamma : [t,t_0] \to \mathcal{M}$ such that $\gamma(t) = (x,t)$, $\gamma(t_0) \in \overline{B(x_0,t_0,s)}$, and $|\operatorname{Rm}(x,t)| = s^{-2} \geq 10^4 r_+^{-2}$. Hence by Φ-pinching, $R(x,t) \geq \frac{1}{100}s^{-2} \geq 100r_+^{-2}$ (cf. the remark just before the statement of Lemma 79.12). With reference to the constant η of (69.2), put $\sigma = 10^{-6} \min\left(1, \frac{1}{\eta}\right)$. Let $\alpha : [0, \widehat{\rho}] \to \overline{B(x_0,t_0,s)} \subset \mathcal{M}_{t_0}^-$ be a minimizing geodesic from (x_0,t_0) to $\gamma(t_0) \in \overline{B(x_0,t_0,s)}$. We can find a point z along α with $\operatorname{dist}_{t_0}(z, \gamma(t_0)) \leq \sigma s$ and $\operatorname{dist}_{t_0}(z, x_0) \leq (1-\sigma)s$. Let $\bar{\gamma} : [t,t_0] \to \mathcal{M}$ be the static curve ending at z and put $(\bar{x},t) = \bar{\gamma}(t)$. In brief, we get $(\bar{x},t)$ by "pulling (x,t) slightly inward from the boundary".

From the distance distortion estimate of Section 27, we can say that $\operatorname{dist}_t(\bar{x},x) \leq 10^6 \sigma s$. Then applying (69.2) along a minimizing time-t curve from $\bar{x}$ to x, we conclude that

$$|R^{-\frac{1}{2}}(\bar{x},t) - R^{-\frac{1}{2}}(x,t)| \leq \frac{1}{2}\eta \operatorname{dist}_t(\bar{x},x) \leq \frac{1}{2}s, \tag{79.24}$$

so $R^{-\frac{1}{2}}(\bar{x},t) \leq R^{-\frac{1}{2}}(x,t) + \frac{1}{2}s \leq 20s$ and hence $R(\bar{x},t) \geq \frac{1}{400}s^{-2} \geq 25 r_+^{-2}$. In particular, $(\bar{x},t)$ has a canonical neighborhood. We also know that $R(\bar{x},t) \leq 6|\operatorname{Rm}(\bar{x},t)| \leq 6s^{-2}$. It follows from the definition of canonical neighborhoods (see Definition 69.1) that there is some universal constant so that $\operatorname{vol}(B(\bar{x},t,10^{-9}s)) \geq \text{const.}\, s^3$. (We recall that from Lemma 60.3, there is a $\kappa > 0$ such that a standard solution is κ-noncollapsed as scales < 1.) The distance distortion estimate ensures that $B(\bar{x},t,10^{-9}s) \subset B(z,t_0,10^{-6}s) \subset B(x_0,t_0,s)$. Then the standard volume distortion estimate implies that $\operatorname{vol}(B(x_0,t_0,s)) \geq \text{const.}\operatorname{vol}(B(\bar{x},t,10^{-9}s)) \geq \text{const.}\, s^3$, again for some universal constant. Finally we use Bishop-Gromov volume comparison to get $\operatorname{vol}(B(x_0,t_0,r_0))r_0^{-3} \geq \text{const.}\operatorname{vol}(B(x_0,t_0,s))s^{-3}$.

In case (c) we apply (79.21), replacing the r_0 parameter there by s, and Bishop-Gromov volume comparison as in case (b). □

80. Construction of the Ricci flow with surgery

The proof is by induction on i. To start the induction process, we observe that the initial normalization $|\operatorname{Rm}| \leq 1$ at $t = 0$ implies that a smooth solution exists for some definite time [22, Corollary 7.7]. The curvature bound on this time interval, along with the volume assumption on the initial time balls, implies that the solution is κ-noncollapsed below scale 1 and satisfies the ρ-canonical neighborhood assumption vacuously for small $\rho > 0$.

Now assume inductively that r_j, κ_j, and $\bar{\delta}_j$ have been selected for $1 \leq j \leq i$, thereby defining the functions r, κ, and $\bar{\delta}$ on $[0, 2^i\epsilon]$, such that for any nonincreasing function δ on $[0, 2^i\epsilon]$ satisfying $0 < \delta(t) \leq \bar{\delta}(t)$, if one has normalized initial data then the Ricci flow with (r,δ)-cutoff is defined on $[0, 2^i\epsilon]$ and is $\kappa(t)$-noncollapsed at scales $< \epsilon$.

We will determine κ_{i+1} using Lemma 79.12. So suppose it is not possible to choose r_{i+1} and $\bar{\delta}_{i+1}$ (and, if necessary, make $\bar{\delta}_i$ smaller) so that if we put $\kappa_{i+1} = \kappa_+(r_i, \kappa_i, 2^{i-1}\epsilon, 3(2^{i-1})\epsilon)$ (where κ_+ denotes the function from Lemma 79.12) then the inductive statement above holds with i replaced by $i+1$. Then given sequences $r^\alpha \to 0$ and $\bar{\delta}^\alpha \to 0$, for each α there must be a counterexample, say $(M^\alpha, g^\alpha(0))$, to the statement with $r_{i+1} = r^\alpha$ and $\bar{\delta}_i = \bar{\delta}_{i+1} = \bar{\delta}^\alpha$. We assume that

$$\bar{\delta}^\alpha < \hat{\delta}(\alpha, 1 - \frac{1}{\alpha}, r^\alpha) \tag{80.1}$$

where $\hat{\delta}$ is the quantity from Lemma 74.1; this will guarantee that for any $A < \infty$ and $\theta \in (0,1)$ we may apply Lemma 74.1 with parameters A and θ for sufficiently large α. We also assume that

$$\bar{\delta}^\alpha < \bar{\delta}(r_i, r^\alpha, \kappa_i, 2^{i-1}\epsilon, 3(2^{i-1})\epsilon) \tag{80.2}$$

where $\bar{\delta}(r_i, r^\alpha, \kappa_i, 2^{i-1}\epsilon, 3(2^{i-1})\epsilon)$ is from Lemma 79.12. By Lemma 73.7 each initial condition $(M^\alpha, g^\alpha(0))$ will prolong to a Ricci flow with surgery $\mathcal{M}^\alpha$ defined on a time interval $[0, T^\alpha]$ with $T^\alpha \in (2^i\epsilon, \infty]$, which restricts to a Ricci flow with (r, δ)-cutoff on any proper subinterval $[0, \tau]$ of $[0, T^\alpha]$, but for which the r^α-canonical neighborhood assumption fails at some point $(\bar{x}^\alpha, T^\alpha)$ lying in the backward time slice $\mathcal{M}^{\alpha-}_{T^\alpha}$. (It is implicit in this statement that $R(\bar{x}^\alpha, T^\alpha) \geq \frac{1}{(r^\alpha)^2}$.) Since $(M^\alpha, g^\alpha(0))$ violates the theorem, we must have $T^\alpha \in (2^i\epsilon, 2^{i+1}\epsilon]$. By (80.2) and Lemma 79.12, it follows that $\mathcal{M}^\alpha$ is κ_{i+1}-noncollapsed at scales below ϵ on the interval $[2^i\epsilon, T^\alpha)$, where $\kappa_{i+1} = \kappa_+(r_i, \kappa_i, 2^{i-1}\epsilon, 3(2^{i-1})\epsilon)$ and κ_+ denotes the function from Lemma 79.12.

Let $(\widehat{\mathcal{M}}^\alpha, (\bar{x}^\alpha, 0))$ be the pointed Ricci flow with surgery obtained from $(\mathcal{M}^\alpha, (\bar{x}^\alpha, T^\alpha))$ by shifting time by T^α and parabolically rescaling by $R(\bar{x}^\alpha, T^\alpha)$. We also remove the part of $(\widehat{\mathcal{M}}^\alpha, (\bar{x}^\alpha, 0))$ after time zero and we take the time-zero slice $\widehat{\mathcal{M}}^\alpha_0$ to be diffeomorphic to $\mathcal{M}^{\alpha-}_{T^\alpha}$. In brief, the rest of the proof goes as follows. If surgeries occur further and further away from $(\bar{x}^\alpha, 0)$ in spacetime as $\alpha \to \infty$, then the reasoning of Theorem 52.7 applies and we obtain a κ-solution as a limit. This would contradict the fact that $(\bar{x}^\alpha, T^\alpha)$ does not have a canonical neighborhood. Thus there must be surgeries in a parabolic ball of a fixed size centered at $(\bar{x}^\alpha, 0)$, for arbitrarily large α. Then one argues using Lemma 74.1 that the solution will be close to the (suitably rescaled and time-shifted) standard solution, which again leads to a canonical neighborhood and a contradiction.

We now return to the proof. Recall that a metric ball B is proper if the distance function from the center is a proper function on B. If T is a surgery time for a Ricci flow with surgery then a metric ball in $\mathcal{M}^-_T$ need not be proper.

Note that by continuity, every point in $\widehat{\mathcal{M}}^\alpha$ whose scalar curvature is strictly greater than that of $(\bar{x}^\alpha, 0)$ has a neighborhood as in Definition 69.1, except that the error estimate is 2ϵ instead of ϵ.

Sublemma 80.3. *For all $\lambda < \infty$, the ball $B(\bar{x}^\alpha, 0, \lambda) \subset \widehat{\mathcal{M}}^\alpha_0$ is proper for sufficiently large α.*

Proof. As in Lemma 70.2, for each $\rho < \infty$ the scalar curvature on $B(\bar{x}^\alpha, 0, \rho) \subset \widehat{\mathcal{M}}^\alpha_0$ is uniformly bounded in terms of α. (In carrrying out the proof of Lemma 70.2, we now use the

aforementioned property of having canonical neighborhoods of quality 2ϵ.) Thus $B(\bar{x}^\alpha, 0, \rho)$ has compact closure in $\widehat{\mathcal{M}}_0^\alpha$ (see Lemma 67.9), from which the sublemma follows. $\square$

Let $T_1 \in [-\infty, 0]$ be the infimum of the set of numbers $\tau' \in (-\infty, 0]$ such that for all $\lambda < \infty$, the ball $B(\bar{x}^\alpha, 0, \lambda) \subset \widehat{\mathcal{M}}_0^\alpha$ is proper, and unscathed on $[\tau', 0]$ for sufficiently large α.

Lemma 80.4. *After passing to a subsequence if necessary, the pointed flows $(\widehat{\mathcal{M}}^\alpha, (\bar{x}^\alpha, 0))$ converge on the time interval $(T_1, 0]$ to a Ricci flow (without surgery) $(\mathcal{M}^\infty, (\bar{x}^\infty, 0))$ with a smooth complete nonnegatively-curved Riemannian metric on each time slice, and scalar curvature globally bounded above by some number $Q < \infty$. (We interpret $(0, 0]$ to mean $\{0\}$ rather than the empty set.)*

Proof. Suppose first that $T_1 < 0$. Then the arguments of Theorem 52.7 apply in the time interval $(T_1, 0]$, to give the Ricci flow (without surgery) $(\mathcal{M}^\infty, (\bar{x}^\infty, 0))$. Since $r^\alpha \to 0$, Hamilton-Ivey pinching implies that $\mathcal{M}^\infty$ will have nonnegative curvature. The fact that the canonical neighborhood assumption, with ϵ replaced by 2ϵ, holds for each $\widehat{\mathcal{M}}^\alpha$ allows us to deduce that the scalar curvature of $\mathcal{M}^\infty$ is globally bounded above by some number $Q < \infty$; compare with Section 46 and Step 4 of the proof of Theorem 52.7.

Now suppose that $T_1 = 0$. The argument is similar to Steps 2 and 3 of the proof of Theorem 52.7. As in Step 2, or more precisely as in Lemma 70.2, for each $\rho < \infty$ the scalar curvature on $B(\bar{x}^\alpha, 0, \rho) \subset \widehat{\mathcal{M}}_0^\alpha$ is uniformly bounded in terms of α. Given $\rho < \infty$ and $(x^\alpha, 0) \in B(\bar{x}^\alpha, 0, \rho) \subset \widehat{\mathcal{M}}^\alpha$, if the parabolic region of Lemma 70.1 (centered around $(x^\alpha, 0) \in \widehat{\mathcal{M}}^\alpha$) is unscathed then we can apply the 2ϵ-canonical neighborhood assumption on $\widehat{\mathcal{M}}^\alpha$, Lemma 70.1 and Appendix D to derive bounds on the curvature derivatives at $(x^\alpha, 0) \in \widehat{\mathcal{M}}_0^\alpha$ that depend on ρ but are independent of α. If the parabolic region is scathed then we can apply Lemma 74.1, along with our scalar curvature bound at $(x^\alpha, 0)$, to again obtain uniform bounds on the curvature derivatives at $(x^\alpha, 0)$. Hence there is a subsequence of the pointed Riemannian manifolds $\{(\widehat{\mathcal{M}}_0^\alpha, \bar{x}^\alpha)\}_{\alpha=1}^\infty$ that converges to a smooth complete pointed Riemannian manifold $(\mathcal{M}_0^\infty, \bar{x}^\infty)$. As in the previous case, it will have bounded nonnegative sectional curvature.

This proves the lemma. Alternatively, in the case $T_1 = 0$ one can argue directly that if the parabolic region of Lemma 70.1 is scathed then $(\overline{x}^\alpha, T^\alpha)$ has a canonical neighborhood; see the rest of the proof of Proposition 77.2. $\square$

If we can show that $T_1 = -\infty$ then $(\mathcal{M}^\infty, (\bar{x}^\infty, 0))$ will be a κ-solution, which will contradict the assumption that $(\bar{x}^\alpha, T^\alpha)$ does not admit a canonical neighborhood. Suppose that $T_1 > -\infty$. We know that for all $\tau' \in (T_1, 0]$ and $\lambda < \infty$, the scalar curvature in $P(\bar{x}^\alpha, 0, \lambda, \tau')$ is bounded by $Q + 1$ when α is sufficiently large. By Lemma 70.1, there exists $\sigma < T_1$ such that for all $\lambda < \infty$, if (for large α) the solution $\widehat{\mathcal{M}}^\alpha$ is unscathed on $P(\bar{x}^\alpha, 0, \lambda, t^\alpha)$ for some $t^\alpha > \sigma$ then

$$R(x,t) < 8(Q+2) \quad \text{for all} \quad (x,t) \in P(\bar{x}^\alpha, 0, \lambda, t^\alpha). \tag{80.5}$$

(In applying Lemma 70.1, we use the fact that in the unscaled variables, $r(T^\alpha)^{-2} \leq R(\bar{x}^\alpha, T^\alpha)$ by assumption, along with the fact that $r(\cdot)$ is a nonincreasing function.)

By the definition of T_1, and after passing to a subsequence if necessary, there exist $\lambda < \infty$ and a sequence $\gamma^\alpha : [\sigma^\alpha, 0] \to \widehat{\mathcal{M}}^\alpha$ of static curves so that
1. $\gamma^\alpha(0) \in B(\bar{x}^\alpha, 0, \lambda)$ and
2. The point $\gamma^\alpha(\sigma^\alpha)$ is inserted during surgery at time $\sigma^\alpha > \sigma$.

For each α, we may assume that σ^α is the largest number having this property. Put $\xi^\alpha = \sigma^\alpha + (h^\alpha(\sigma^\alpha))^2$. (In the notation of [52], $(h^\alpha(\sigma^\alpha))^2$ would be written as $R(\bar{x}, \bar{t})\, h^2(T_0)$. Note that we have no *a priori* control on $h^\alpha(\sigma^\alpha)$.) Then ξ^α is the blowup time of the rescaled and shifted standard solution that Lemma 74.1 compares with $(\widehat{\mathcal{M}}^\alpha, \gamma^\alpha(\sigma^\alpha))$. We claim that $\liminf_{\alpha \to \infty} \xi^\alpha > 0$. Otherwise, Lemma 74.1 would imply that after passing to a subsequence, there are regions of $\widehat{\mathcal{M}}^\alpha$, starting from time σ^α, that are better and better approximated by rescaled and shifted standard solutions whose blowup times ξ^α have a limit that is nonpositive, thereby contradicting (80.5). Lemma 74.1, along with the fact that $R(\bar{x}^\alpha, 0) = 1$, also gives a uniform upper bound on ξ^α.

Now Lemma 74.1 implies that for large α, the restriction of $\widehat{\mathcal{M}}^\alpha$ to the time interval $[\sigma^\alpha, 0]$ is well approximated by the restriction to $[\sigma^\alpha, 0]$ of a rescaled and shifted standard solution. Then Lemma 63.1 implies that $(\bar{x}^\alpha, T^\alpha)$ has a canonical neighborhood. The canonical neighborhood may be either a strong ϵ-neck or an ϵ-cap. (Note a strong ϵ-neck may arise when an ϵ-neck around $(\bar{x}^\alpha, T^\alpha)$ extends smoothly backward in time to form a strong ϵ-neck that incorporates part of the Ricci flow solution that existed before the surgery time σ^α.)

This is a contradiction. □

81. II.6. Double sided curvature bound in the thick part

Having shown that for a suitable choice of the functions r and δ, the Ricci flow with (r, δ)-cutoff exists for all time and for every normalized initial condition, one wants to understand its implications. The main results in II.6 are noncollapsing and curvature estimates which form the basis of the analysis of the large-time behavior given in II.7.

Lemma 81.1. *If $\mathcal{M}$ is a Ricci flow with (r, δ)-cutoff on a compact manifold and $g(0)$ has positive scalar curvature then the solution goes extinct after a finite time, i.e. $\mathcal{M}_T = \emptyset$ for some $T > 0$.*

Proof. We apply (B.2). This formula is initially derived for smooth flows but because surgeries are performed in regions of high scalar curvature, it is also valid for a Ricci flow with surgery; cf. the proof of Lemma 79.11. It follows that the flow goes extinct by time $\frac{3}{2R_{min}(0)}$. □

Lemma 81.2. *If $\mathcal{M}$ is a Ricci flow with surgery that goes extinct after a finite time, then the initial (compact connected orientable) 3-manifold is diffeomorphic to a connected sum of $S^1 \times S^2$'s and quotients of the round S^3.*

Proof. This follows from Lemma 73.4. □

According to [24, 25] and [53], if none of the prime factors in the Kneser-Milnor decomposition of the initial manifold are aspherical then the Ricci flow with surgery again goes extinct after a finite time. Along with Lemma 81.2, this proves the Poincaré Conjecture.

Passing to Ricci flow solutions that may not go extinct after a finite time, the main result of II.6 is the following :

Corollary 81.3. *(cf. Corollary II.6.8) For any $w > 0$ one can find $\tau = \tau(w) > 0$, $K = K(w) > 0$, $\overline{r} = \overline{r}(w) > 0$ and $\theta = \theta(w) > 0$ with the following property. Suppose we have a solution to the Ricci flow with (r, δ)-cutoff on the time interval $[0, t_0]$, with normalized initial data. Let $h_{\max}(t_0)$ be the maximal surgery radius on $[t_0/2, t_0]$. (If there are no surgeries on $[t_0/2, t_0]$ then $h_{\max}(t_0) = 0$.) Let r_0 satisfy*
1. $\theta^{-1}(w) h_{\max}(t_0) \le r_0 \le \overline{r}\sqrt{t_0}$.
2. The ball $B(x_0, t_0, r_0)$ has sectional curvatures at least $-r_0^{-2}$ at each point.
3. $\mathrm{vol}(B(x_0, t_0, r_0)) \ge w r_0^3$.

Then the solution is unscathed in $P(x_0, t_0, r_0/4, -\tau r_0^2)$ and satisfies $R < K r_0^{-2}$ there.

Corollary 81.3 is an analog of Corollary 55.1, but there are some differences. One minor difference is that Corollary 55.1 is stated as the contrapositive of Corollary 81.3. Namely, Corollary 55.1 assumes that $-r_0^{-2}$ is achieved as a sectional curvature in $B(x_0, t_0, r_0)$, and its conclusion is that $\mathrm{vol}(B(x_0, t_0, r_0)) \le w r_0^3$. The relation with Corollary 81.3 is the following. Suppose that assumptions 1 and 2 of Corollary 81.3 hold. If $-r_0^{-2}$ is achieved somewhere as a sectional curvature in $B(x_0, t_0, r_0)$ then Hamilton-Ivey pinching implies that the scalar curvature is very large at that point, which contradicts the conclusion of Corollary 81.3. Hence assumption 3 of Corollary 81.3 must not be satisfied.

A more substantial difference is that the smoothness of the flow in Corollary 55.1 is guaranteed by the setup, whereas in Corollary 81.3 we must prove that the solution is unscathed in $P(x_0, t_0, r_0/4, -\tau r_0^2)$.

The role of the parameter $\overline{r}$ in Corollary 81.3 is essentially to guarantee that we can use Hamilton-Ivey pinching effectively.

82. II.6.5. Earlier scalar curvature bounds on smaller balls from lower curvature bounds and a later volume bound

A simpler analog of Corollary 81.3 is the following.

Lemma 82.1. *(cf. Lemma II.6.5(a)) Given $w > 0$, there exist $\tau_0 = \tau_0(w) > 0$ and $K_0 = K_0(w) < \infty$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff such that*
1. The parabolic neighborhood $P(x_0, 0, r_0, -\tau r_0^2)$ is unscathed, where $\tau \le \tau_0$.
2. The sectional curvatures are bounded below by $-r_0^{-2}$ on $P(x_0, 0, r_0, -\tau r_0^2)$.
3. $\mathrm{vol}(B(x_0, 0, r_0)) \ge w r_0^3$.

Then $R \le K_0 \, \tau^{-1} \, r_0^{-2}$ on $P(x_0, 0, r_0/4, -\tau r_0^2/2)$.

Proof. Let us first note a consequence of Corollary 45.13.

Sublemma 82.2. *Given $w > 0$, there exist $\tau_0 = \tau_0(w) > 0$ and $K = K(w) < \infty$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff such that*
1. There are no surgeries on a family $\bigcup_{t \in [-\tau r_0^2, 0]} B(x_0, t, r_0)$ of time-dependent balls, where $\tau \leq \tau_0$.
2. The sectional curvatures are bounded below by $- r_0^{-2}$ on the above family of balls.
3. $\mathrm{vol}(B(x_0, 0, r_0)) \geq w r_0^3$.

Then $R \leq K \tau^{-1} r_0^{-2}$ on $\bigcup_{t \in [-\frac{3}{4}\tau r_0^2, 0]} B(x_0, t, r_0/2)$.

Here we have changed the conclusion of Corollary 45.13 to obtain an upper curvature bound on $\bigcup_{t \in [-\frac{3}{4}\tau r_0^2, 0]} B(x_0, t, r_0/2)$ instead of $\bigcup_{t \in [-\frac{3}{4}\tau r_0^2, 0]} B(x_0, t, r_0/4)$, but this clearly follows from the arguments of the proof of Corollary 45.13.

We now return to the situation of Lemma 82.1.

If τ_0 is sufficiently small, then for $t \in [-\tau r_0^2, 0]$ and $(x, t) \in B(x_0, t, 9r_0/10)$, the lower curvature bound $\mathrm{Rm} \geq - r_0^{-2}$ on $P(x_0, 0, r_0, -\tau r_0^2)$ implies that $(x, 0) \in B(x_0, 0, r_0)$ (more precisely, that (x, t) lies on a static curve with one endpoint in $B(x_0, 0, r_0)$, or equivalently, that $(x, t) \in P(x_0, 0, r_0, -\tau r_0^2)$). Thus $\bigcup_{t \in [-\tau r_0^2, 0]} B(x_0, t, 9r_0/10) \subset P(x_0, 0, r_0, -\tau r_0^2)$ and so $\mathrm{Rm} \geq -r_0^{-2} \geq -(9r_0/10)^{-2}$ on $\bigcup_{t \in [-\tau (9r_0/10)^2, 0]} B(x_0, t, 9r_0/10)$.

Applying Sublemma 82.2 with r_0 replaced by $9r_0/10$, and slightly redefining w, gives that $R \leq K \tau^{-1} (9r_0/10)^{-2}$ on $\bigcup_{t \in [-\frac{3}{4}\tau (9r_0/10)^2, 0]} B(x_0, t, 9r_0/20)$. Then the length distortion estimate of Lemma 27.8 implies that for sufficiently small τ_0, if $(x, 0) \in B(x_0, 0, r_0/4)$ then $(x, t) \in B(x_0, t, 9r_0/20)$ for $t \in [-\tau r_0^2/2, 0]$. That is,

$$P(x_0, 0, r_0/4, -\tau r_0^2/2) \subset \bigcup_{t \in [-\tau r_0^2/2, 0]} B(x_0, t, 9r_0/20). \tag{82.3}$$

In applying the length distortion estimate we use the fact that the change in distance is estimated by $\Delta d \leq \text{const.} \sqrt{K \tau^{-1} (9r_0/10)^{-2}} \cdot \tau r_0^2/2$ which, for small τ_0, is a small fraction of r_0.

Thus we have shown that $R \leq K \tau^{-1} (9r_0/10)^{-2}$ on $P(x_0, 0, r_0/4, -\tau r_0^2/2)$. This proves the lemma. □

The formulation of [52, Lemma II.6.5] specializes Lemma 82.1 to the case $w = 1 - \epsilon$. It includes the statement [52, Lemma II.6.5(b)] saying that $\mathrm{vol}(B(x_0, r_0/4, -\tau r_0^2))$ is at least $\frac{1}{10}$ of the volume of the Euclidean ball of the same radius. This follows from the proof of Corollary 45.1(b), provided that τ_0 is sufficiently small.

There is an evident analogy between Lemma 82.1 and Corollary 81.3. However, there is the important difference that Corollary 81.3 (along with Corollary 55.1) only assumes a lower sectional curvature bound at the final time slice.

83. II.6.6. Locating small balls whose subballs have almost Euclidean volume

The result of this section is a technical lemma about volumes of subballs.

Lemma 83.1. *(cf. Lemma II.6.6) For any $\widehat{\epsilon}, w > 0$ there exists $\theta_0 = \theta_0(\widehat{\epsilon}, w)$ such that if $B(x,1)$ is a metric ball of volume at least w, compactly contained in a manifold without boundary with sectional curvatures at least -1, then there exists a subball $B(y,\theta_0) \subset B(x,1)$ such that every subball $B(z,r) \subset B(y,\theta_0)$ of any radius has volume at least $(1-\widehat{\epsilon})$ times the volume of the Euclidean ball of the same radius.*

The proof is similar to that of Lemma 54.1. Suppose that the claim is not true. Then there is a sequence of Riemannian manifolds $\{M_i\}_{i=1}^{\infty}$ and balls $B(x_i,1) \subset M_i$ with compact closure so that $\mathrm{Rm}\Big|_{B(x_i,1)} \geq -1$ and $\mathrm{vol}(B(x_i,1)) \geq w$, along with a sequence $r_i' \to 0$ so that each subball $B(x_i', r_i') \subset B(x_i,1)$ has a subball $B(x_i'', r_i'') \subset B(x_i', r_i')$ with $\mathrm{vol}(B(x_i'', r_i'')) < (1 - \widehat{\epsilon})\, \omega_3 (r'')^n$. After taking a subsequence, we can assume that $\lim_{i\to\infty}(B(x_i,1),x_i) = (X, x_\infty)$ in the pointed Gromov-Hausdorff topology, where (X, x_∞) is a pointed Alexandrov space with curvature bounded below by -1. From [12, Theorem 10.8], the Riemannian volume forms dvol_{M_i} converge weakly to the three-dimensional Hausdorff measure μ of X. If x'_∞ is a regular point of X then there is some $\delta > 0$ so that $B(x'_\infty, \delta)$ has compact closure in X and for all $r < \delta$, $\mu(B(x'_\infty, r)) \geq (1 - \frac{\widehat{\epsilon}}{10})\omega_3 r^3$. Fixing such an r for the moment, for large i there are balls $B(x_i', r) \subset B(x_i, 1)$ with $\mathrm{vol}(B(x_i', r)) \geq (1 - \frac{\widehat{\epsilon}}{5})\, \omega_3 r^3$. Recalling the sequence $\{r_i'\}$, by hypothesis there is a subball $B(x_i'', r_i'') \subset B(x_i', r_i')$ with $\mathrm{vol}(B(x_i'', r_i'')) < (1 - \widehat{\epsilon})\, \omega_3 (r_i'')^3$. Clearly $B(x_i', r) \subset B(x_i'', r + r_i')$. From the Bishop-Gromov inequality,

$$\frac{\mathrm{vol}(B(x_i'', r + r_i'))}{\mathrm{vol}(B(x_i'', r_i''))} \leq \frac{\int_0^{r+r_i'} \sinh^2(s)\, ds}{\int_0^{r_i''} \sinh^2(s)\, ds}. \tag{83.2}$$

Then

$$\mathrm{vol}(B(x_i', r)) \leq \mathrm{vol}(B(x_i'', r + r_i')) \leq (1 - \widehat{\epsilon})\, \omega_3 \frac{(r_i'')^3}{\int_0^{r_i''} \sinh^2(s)\, ds} \int_0^{r+r_i'} \sinh^2(s)\, ds. \tag{83.3}$$

For large i we obtain

$$\mathrm{vol}(B(x_i', r)) \leq (1 - \frac{\widehat{\epsilon}}{2})\, \omega_3 \cdot 3 \int_0^r \sinh^2(s)\, ds. \tag{83.4}$$

Then if we choose r to be sufficiently small, we contradict the fact that $\mathrm{vol}(B(x_i', r)) \geq (1 - \frac{\widehat{\epsilon}}{5})\, \omega_3 r^3$ for all i.

Remark 83.5. By similar reasoning, for every $L > 1$ one may find $\theta_1 = \theta_1(\widehat{\epsilon}, L)$ such that under the hypotheses of Lemma 83.1, there is a subball $B(y, \theta_1) \subset B(x,1)$ which is L-biLipschitz to the Euclidean unit ball.

84. II.6.8. Proof of the double sided curvature bound in the thick part, modulo two propositions

In this section we explain how Corollary 81.3 follows from Lemma 83.1 and two other propositions, which will be proved in subsequent sections. We first state the other propositions, which are Propositions 84.1 and 84.2.

Proposition 84.1. *(cf. Proposition II.6.3) For any $A < \infty$ one can find positive constants $\kappa(A)$, $K_1(A)$, $K_2(A)$, $\overline{r}(A)$, such that for any $t_0 < \infty$ there exists $\overline{\delta}_A(t_0) > 0$, decreasing in t_0, with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff on a time interval $[0, T]$, where $\delta(t) < \overline{\delta}_A(t)$ on $[t_0/2, t_0]$, with normalized initial data. Assume that*
1. The solution is unscathed on a parabolic ball $P(x_0, t_0, r_0, -r_0^2)$, with $2r_0^2 < t_0$.
2. $|\operatorname{Rm}| \leq \frac{1}{3r_0^2}$ on $P(x_0, t_0, r_0, -r_0^2)$.
3. $\operatorname{vol}(B(x_0, t_0, r_0)) \geq A^{-1} r_0^3$.

Then
(a) The solution is κ-noncollapsed on scales less than r_0 in $B(x_0, t_0, Ar_0)$.
(b) Every point $x \in B(x_0, t_0, Ar_0)$ with $R(x, t_0) \geq K_1 r_0^{-2}$ has a canonical neighborhood in the sense of Definition 69.1.
(c) If $r_0 \leq \overline{r}\sqrt{t_0}$ then $R \leq K_2 r_0^{-2}$ in $B(x_0, t_0, Ar_0)$.

Proposition 84.1(a) is an analog of Theorem 28.2.

(The reason for the "3" in the hypothesis $|\operatorname{Rm}| \leq \frac{1}{3r_0^2}$ comes from Remark 28.3.) Proposition 84.1(c) is an analog of Theorem 53.1, but the hypotheses are slightly different. In Proposition 84.1 one assumes a lower bound on the volume of the time-t_0 ball $B(x_0, t_0, r_0)$, while in Theorem 53.1 one assumes assumes a lower bound on the volume of the time-$(t_0 - r_0^2)$ ball $B(x_0, t_0 - r_0^2, r_0)$. In view of the curvature assumption on $P(x_0, t_0, r_0, -r_0^2)$, the hypotheses are essentially equivalent.

Conclusions (a), (b) and (c) of Proposition 84.1 are similar to the conclusions of Theorem 28.2, Lemma 53.3 and Theorem 53.1, respectively. Conclusions (a) and (b) of Proposition 84.1 are also related to what was proved in Proposition 77.2 to construct the Ricci flow with surgery. The difference is that the noncollapsing and canonical neighborhood results of Proposition 77.2 are statements at or below the scale $r(t)$, whereas Proposition 84.1 is a statement about much larger scales, comparable to $\sqrt{t_0}$. We note that the parameter $\overline{\delta}_A$ in Proposition 84.1 is independent of the function δ used to define the Ricci flow with (r, δ)-cutoff.

In the proof of the next proposition we will apply Lemma 83.1 with $\widehat{\epsilon}$ equal to the global parameter ϵ, so we will write $\theta(w)$ instead of $\theta(\epsilon, w)$.

Proposition 84.2. *(cf. Proposition II.6.4) There exist $\tau, \overline{r}, C_1 > 0$ and $K < \infty$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff on the time interval $[0, t_0]$, with normalized initial data. Let r_0 satisfy $2C_1 h_{\max}(t_0) \leq r_0 \leq \overline{r}\sqrt{t_0}$, where $h_{\max}(t_0)$ is the maximal cutoff radius for surgeries in $[t_0/2, t_0]$. (If there are no surgeries on $[t_0/2, t_0]$ then $h_{\max}(t_0) = 0$.)*

Assume
1. The ball $B(x_0, t_0, r_0)$ has sectional curvatures at least $-r_0^{-2}$ at each point.
2. The volume of any subball $B(x, t_0, r) \subset B(x_0, t_0, r_0)$ with any radius $r > 0$ is at least $(1-\epsilon)$ times the volume of the Euclidean ball of the same radius.

Then the solution is unscathed on $P(x_0, t_0, r_0/4, -\tau r_0^2)$ and satisfies $R < K r_0^{-2}$ there.

Proposition 84.2 is an analog of Theorem 54.2. However, there is the important difference that in Proposition 84.2 we have to prove that no surgeries occur within $P(x_0, t_0, r_0/4, -\tau r_0^2)$.

Assuming the validity of Propositions 84.1 and 84.2, suppose that the hypotheses of Corollary 81.3 are satisfied. We will allow ourselves to shrink the parameter $\overline{r}$ in order to apply Hamilton-Ivey pinching when needed. Put $r_0' = \theta_0(w)\, r_0$, where $\theta_0(w)$ is from Lemma 83.1. By Lemma 83.1, there is a subball $B(x_0', t_0, r_0') \subset B(x_0, t_0, r_0)$ such that every subball of $B(x_0', t_0, r_0')$ has volume at least $(1-\epsilon)$ times the volume of the Euclidean ball of the same radius. As the sectional curvatures are bounded below by $-r_0^{-2}$ on $B(x_0, t_0, r_0)$, they are bounded below by $-(r_0')^{-2}$ on $B(x_0', t_0, r_0')$. By an appropriate choice of the parameters $\theta(w)$ and $\overline{r}$ of Corollary 81.3, in particular taking $\theta(w) \le \frac{\theta_0(w)}{2C_1}$, we can ensure that Proposition 84.2 applies to $B(x_0', t_0, r_0')$. Then the solution is unscathed on $P(x_0', t_0, r_0'/4, -\tau (r_0')^2)$ and satisfies $|\operatorname{Rm}| \le K\,(r_0')^{-2}$ there, where the lower bound on Rm comes from Hamilton-Ivey pinching. With τ being the parameter of Proposition 84.2 and putting $r_0'' = \min(K^{-1/2}, \tau^{1/2}, \frac{1}{4})\, r_0'$, for all $t_0'' \in [t_0 - (r_0'')^2, t_0]$ the solution is unscathed on $P(x_0', t_0'', r_0'', -(r_0'')^2)$ and satisfies $|\operatorname{Rm}| \le (r_0'')^{-2}$ there. From the curvature bound $\operatorname{Rm} \ge -(r_0')^{-2}$ on $P(x_0', t_0, r_0'/4, -\tau(r_0')^2)$ (coming from pinching) and the fact that $B(x_0', t_0, r_0'')$ has almost Euclidean volume, we obtain a bound $\operatorname{vol}(B(x_0', t_0'', r_0'')) \ge \text{const.}\,(r_0'')^3$. Applying Proposition 84.1 with $A = \frac{100 r_0}{r_0''}$ gives $R \le K_2 (r_0'')^{-2}$ on $B(x_0, t_0'', 10 r_0) \subset B(x_0', t_0'', 100 r_0)$, for all $t_0'' \in [t_0 - (r_0'')^2, t_0]$. Writing this as $R \le \text{const.}\, r_0^{-2}$, if we further restrict $\theta(w)$ to be sufficiently small then we can ensure that $R \le \text{const.}\ \theta^2(w)\, h^{-2} \le .01\, h^{-2}$. As surgeries only occur at spacetime points (x,t) where $R(x,t) \sim h(t)^{-2}$, there are no surgeries on $\bigcup_{t_0'' \in [t_0 - (r_0'')^2, t_0]} B(x_0, t_0'', 10 r_0)$. Using length distortion estimates, we can find a parabolic neighborhood $P(x_0, t_0, r_0/4, -\tau r_0^2) \subset \bigcup_{t_0'' \in [t_0 - (r_0'')^2, t_0]} B(x_0, t_0'', 10 r_0)$ for some fixed τ. This proves Corollary 81.3.

85. II.6.3. Canonical neighborhoods and later curvature bounds on bigger balls from curvature and volume bounds

We now prove Proposition 84.1. We first recall its statement.

Proposition 85.1. *(cf. Proposition II.6.3) For any $A > 0$ one can find positive constants $\kappa(A)$, $K_1(A)$, $K_2(A)$, $\overline{r}(A)$, such that for any $t_0 < \infty$ there exists $\overline{\delta}_A(t_0) > 0$, decreasing in t_0, with the following property. Suppose that we have a Ricci flow with (r,δ)-cutoff on a time interval $[0,T]$, where $\delta(t) < \overline{\delta}_A(t)$ on $[t_0/2, t_0]$, with normalized initial data. Assume that*
1. The solution is unscathed on a parabolic neighborhood $P(x_0, t_0, r_0, -r_0^2)$, with $2r_0^2 < t_0$.
2. $|\operatorname{Rm}| \le r_0^{-2}$ on $P(x_0, t_0, r_0, -r_0^2)$.

3. $\mathrm{vol}(B(x_0,t_0,r_0)) \geq A^{-1} r_0^3$.

Then
(a) The solution is κ-noncollapsed on scales less than r_0 in $B(x_0,t_0,Ar_0)$.
(b) Every point $x \in B(x_0,t_0,Ar_0)$ with $R(x,t_0) \geq K_1 r_0^{-2}$ has a canonical neighborhood in the sense of Definition 69.1.
(c) If $r_0 \leq \overline{r}\sqrt{t_0}$ then $R \leq K_2 r_0^{-2}$ in $B(x_0,t_0,Ar_0)$.

Proof. The proof of part (a) is analogous to the proof of Theorem 28.2. The proof of part (b) is analogous to the proof of Lemma 53.3. The proof of part (c) is analogous to the proof of Theorem 53.1. We will be brief on the parts of the proof of Proposition 84.1 that are along the same lines as was done before, and will concentrate on the differences.

For part (a) : we can reduce the case $r_0 < \frac{r(t_0)}{100}$ to the case $r_0 \geq \frac{r(t_0)}{100}$ as in Sublemma 79.23; a canonical neighborhood of type (d) with small volume cannot occur, in view of condition 3 of Proposition 84.1. (We remark that the κ-noncollapsing that we want does not follow from the noncollapsing estimate used in the proof of Proposition 77.2, which would give a time-dependent κ.) So we may assume that $\frac{r(t_0)}{100} \leq r_0 \leq \sqrt{t_0/2}$. For fixed t_0, this sets a lower bound on r_0.

Now suppose that $(x,t_0) \in B(x_0,t_0,Ar_0)$, $\rho < r_0$, the parabolic neighborhood $P(x,t_0,\rho,-\rho^2)$ is unscathed and $|\operatorname{Rm}| \leq \rho^{-2}$ there. We want to get a lower bound on $\rho^{-3}\ \mathrm{vol}(B(x,t_0,\rho))$. We recall the idea of the proof of Theorem 28.2. With the notation of Theorem 28.2, after rescaling so that $r_0 = t_0 = 1$, we had a point $x \in B(x_0,1,A)$ around which we wanted to prove noncollapsing. Defining l using curves starting at $(x,1)$, we wanted to find a point $(y,1/2) \in B(x_0,1/2,1/2)$ so that $l(y,1/2)$ was bounded above by a universal constant. Given such a point, we concatenated a minimizing $\mathcal{L}$-geodesic (from $(x,1)$ to $(y,1/2)$) with curves emanating backward in time from $(y,1/2)$. Then the bounded geometry near $(y,1/2)$ allowed us to estimate from below the reduced volume at a time slightly less than $1/2$.

We knew that there was some point $y \in M$ so that $l(y,1/2) \leq \frac{3}{2}$, but the issue in Theorem 28.2 was to find a point $(y,1/2) \in B(x_0,1/2,1/2)$ with $l(y,1/2)$ bounded above by a universal constant. The idea was to take the proof that some point $y \in M$ has $l(y,1/2) \leq \frac{3}{2}$ and localize it near x_0. The proof of Theorem 28.2 used the function $h(y,t) = \phi(d(y,t) - A(2t-1))(\overline{L}(y,1-t)+7)$. Here ϕ was a certain nondecreasing function that is one on $(-\infty,1/20)$ and infinite on $[1/10,\infty)$, and $\overline{L}(q,\tau) = 2\sqrt{\tau}\,L(q,\tau)$. Clearly $\min h(\cdot,1) \leq 7$ and $\min h(\cdot,1/2)$ is achieved in $B(x_0,1/2,1/10)$. The equation $\Box h \geq -(6+C(A))h$ implied that $\frac{d}{dt}\min h \geq -(6+C(A))\min h$, and so $(\min h)(t) \leq 7\, e^{(6+C(A))(1-t)}$.

In the present case, if one knew that the possible contribution of a barely admissible curve to $h(y,t)$ was greater than $7\, e^{(6+C(A))(1-t)} + \epsilon$ then one could still apply the maximum principle to find a point $(y,1/2)$ with $h(y,1/2) \leq 7\, e^{(6+C(A))/2}$. For this, it suffices to know that the possible contribution of a barely admissible curve to $\overline{L}(q,\tau)$ can be bounded below by a sufficiently large number. However, Lemma 79.3 only says that we can make the contribution of a barely admissible curve to L large (using the lower scalar curvature bound to pass from $\mathcal{L}_+$ to $\mathcal{L}$). Because of the factor $2\sqrt{\tau}$ in the definition of $\overline{L}(q,\tau)$, we cannot necessarily say that its contribution to $\overline{L}(q,\tau)$ is large. To salvage the argument, the idea

is to redefine h and redo the proof of Theorem 28.2 in order to get an extra factor of $\sqrt{\tau}$ in $\min h$.

(The use of Lemma 79.3 is similar to what was done in the proof of Proposition 77.2. However, there is a difference in scales. In Proposition 77.2 one was working at a microscopic scale in order to construct the Ricci flow with surgery. The function $\overline{\delta}(t)$ in Proposition 77.2 was relevant to this scale. In the present case we are working at the macroscopic scale $r_0 \sim \sqrt{t_0}$ in order to analyze the long-time behavior of the Ricci flow with surgery. The function $\delta_A(t)$ of Proposition 84.1 is relevant to this scale. Thus we will end up further reducing the surgery function $\overline{\delta}(t)$ of Proposition 77.2 in order to be able to apply Proposition 84.1.)

By assumption, $|\operatorname{Rm}| \leq 1$ at $t = 0$. From Lemma 79.11, $R \geq -\frac{3}{2}\frac{1}{t+1/4}$. Then for $t \in [t_0 - r_0^2/2, t_0]$,

$$R\, r_0^2 \geq -\frac{3}{2}\frac{r_0^2}{t_0 - r_0^2/2} \geq -\frac{3}{2}\frac{r_0^2}{2r_0^2 - r_0^2/2} = -1. \tag{85.2}$$

After rescaling so that $r_0 = 1$, the time interval $[t_0 - r_0^2, t_0]$ is shifted to $[0, 1]$. Then for $t \in [\frac{1}{2}, 1]$, we certainly have $R \geq -3$.

From this, if $0 < \tau \leq \frac{1}{2}$ then $\overline{L}(y,\tau) \geq -6\sqrt{\tau}\int_0^\tau \sqrt{v}\, dv = -4\tau^2$, so $\hat{L}(y,\tau) \equiv \overline{L}(y,\tau) + 2\sqrt{\tau} > 0$.

Putting

$$h(y,\tau) = \phi(d_t(x_0,y) - A(2t-1))\, \hat{L}(y,\tau) \tag{85.3}$$

and using the fact that $\frac{d}{dt}\sqrt{\tau} = -\frac{d}{d\tau}\sqrt{\tau} = -\frac{1}{2\sqrt{\tau}}$, the computations of the proof of Theorem 28.2 give

$$\begin{aligned} \Box h &\geq -\left(\overline{L} + 2\sqrt{\tau}\right) C(A)\,\phi - 6\,\phi - \frac{1}{\sqrt{\tau}}\,\phi \\ &= -C(A)h - \left(6 + \frac{1}{\sqrt{\tau}}\right)\phi. \end{aligned} \tag{85.4}$$

Then if $h_0(\tau) = \min h(\cdot,\tau)$, we have

$$\begin{aligned} \frac{d}{d\tau}\left(\log\left(\frac{h_0(\tau)}{\sqrt{\tau}}\right)\right) &= h_0^{-1}\frac{dh_0}{d\tau} - \frac{1}{2\tau} \leq C(A) + \left(6 + \frac{1}{\sqrt{\tau}}\right)\frac{\phi}{h_0} - \frac{1}{2\tau} \\ &= C(A) + \left(6 + \frac{1}{\sqrt{\tau}}\right)\frac{1}{\overline{L} + 2\sqrt{\tau}} - \frac{1}{2\tau} \\ &= C(A) + \frac{6\sqrt{\tau} + 1}{\sqrt{\tau}\overline{L} + 2\tau} - \frac{1}{2\tau}. \end{aligned} \tag{85.5}$$

As $\overline{L} \geq -4\tau^2$,

$$\frac{d}{d\tau}\left(\log\left(\frac{h_0(\tau)}{\sqrt{\tau}}\right)\right) \leq C(A) + \frac{6\sqrt{\tau} + 1}{2\tau - 4\tau^2\sqrt{\tau}} - \frac{1}{2\tau} \leq C(A) + \frac{50}{\sqrt{\tau}}. \tag{85.6}$$

As $\tau \rightarrow 0$, the Euclidean space computation gives $\overline{L}(q,\tau) \sim |q|^2$, so $\lim_{\tau \rightarrow 0} \frac{h_0(\tau)}{\sqrt{\tau}} = 2$. Then

$$h_0(\tau) \leq 2\sqrt{\tau}\exp(C(A)\tau + 100\sqrt{\tau}). \tag{85.7}$$

This estimate has the desired extra factor of $\sqrt{\tau}$.

It now suffices to show that for a barely admissible curve γ that hits a surgery region at time $1-\tau$,

$$\int_0^\tau \sqrt{v}\,\left(R(\gamma(1-v),v)+|\dot{\gamma}(v)|^2\right)\,dv \geq \exp(C(A)\tau + 100\sqrt{\tau}) + \epsilon, \tag{85.8}$$

where $0<\tau\leq\frac{1}{2}$. Choosing $\overline{\delta}_A(t_0)$ small enough, this follows from Lemma 79.3 along with the lower scalar curvature bound. Then we can apply the maximum principle and follow the proof of Theorem 28.2. In our case the bounded geometry near $(y,1/2)\in B(x_0,1/2,1/2)$ comes from assumptions 2 and 3 of the Proposition. The function $\overline{\delta}_A$ is now determined. After reintroducing the scale r_0, this proves part (a) of the proposition.

The proof of part (b) is similar to the proofs of Lemma 53.3 and Proposition 77.2. Suppose that for some $A>0$ the claim is not true. Then there is a sequence of Ricci flows $\mathcal{M}^\alpha$ which together provide a counterexample. In particular, some point $(x^\alpha,t^\alpha)\in B(x_0^\alpha,t_0^\alpha,Ar_0^\alpha)$ has $R(x^\alpha,t^\alpha)\geq K_1^\alpha(r_0^\alpha)^{-2}$ but does not have a canonical neighborhood, where $K_1^\alpha\rightarrow\infty$ as $\alpha\rightarrow\infty$. Because of the canonical neighborhood assumption, we must have $K_1^\alpha(r_0^\alpha)^{-2}\leq r(t_0^\alpha)^{-2}$. Then $2K_1^\alpha \leq K_1^\alpha\, t_0^\alpha(r_0^\alpha)^{-2} \leq t_0^\alpha\, r(t_0^\alpha)^{-2}$. Since $K_1^\alpha\rightarrow\infty$ and the function $t\rightarrow tr(t)^{-2}$ is bounded on any finite t-interval, it follows that $t_0^\alpha\rightarrow\infty$. Applying point selection to each $\mathcal{M}^\alpha$ and removing the superscripts, there are points $\overline{x}\in B(x_0,\overline{t},2Ar_0)$ with $\overline{t}\in[t_0-r_0^2/2,t_0]$ such that $\overline{Q}\equiv R(\overline{x},\overline{t}) \geq K_1r_0^{-2}$ and $(\overline{x},\overline{t})$ does not have a canonical neighborhood, but each point $(x,t)\in\overline{P}$ with $R(x,t) \geq 4\overline{Q}$ does have a canonical neighborhood, where $\overline{P}=\{(x,t):d_t(x_0,x)\leq d_{\overline{t}}(x_0,\overline{x})+K_1^{1/2}\overline{Q}^{-1/2}, t\in[\overline{t}-\frac{1}{4}K_1\overline{Q}^{-1},\overline{t}]\}$. From (a), we have noncollapsing in $\overline{P}$. Rescaling by $\overline{Q}^{-1}$, we have bounded curvature at bounded distances from $\overline{x}$; see Lemma 70.2. Then we can extract a pointed limit X_∞, which we think of as a time zero slice, that will have nonnegative sectional curvature. (The required pinching for the last statement comes from the assumption that $2r_0^2<t_0$, along with the fact that $K_1^\alpha\rightarrow\infty$.) The fact that points $(x,t)\in\overline{P}$ with $R(x,t) \geq 4\overline{Q}$ have a canonical neighborhood implies that regions of large scalar curvature in X_∞ have canonical neighborhoods, from which one can deduce as in Section 46 that the sectional curvatures of X_∞ are globally bounded above by some $Q_0>0$. Then for each $\overline{A}$, Lemmas 27.8 and 70.1 imply that for large α, the parabolic neighborhood $P(\overline{x},\overline{t},\overline{A}\,\overline{Q}^{-1/2},-\epsilon\eta^{-1}Q_0^{-1}\overline{Q}^{-1})$ is contained in $\overline{P}$. (Here ϵ is a small parameter, which we absorb in the global parameter ϵ.) In applying Lemma 27.8 we use the curvature bound near x_0 coming from the hypothesis of the proposition along with the curvature bound near $\overline{x}$ just derived; cf. the proof of Lemma 53.3. In addition, we claim that $P(\overline{x},\overline{t},\overline{A}\,\overline{Q}^{-1/2},-\epsilon\eta^{-1}Q_0^{-1}\overline{Q}^{-1})$ is unscathed. This is proved as in Section 80. Recall that the idea is to show that a surgery in $P(\overline{x},\overline{t},\overline{A}\,\overline{Q}^{-1/2},-\epsilon\eta^{-1}Q_0^{-1}\overline{Q}^{-1})$ implies that $(\overline{x},\overline{t})$ lies in a canonical neighborhood, which contradicts our assumption. In the argument we use the fact that $t_0^\alpha\rightarrow\infty$ implies $\delta(t_0^\alpha)\rightarrow 0$ in order to rule out surgeries; this is the replacement for the condition $\overline{\delta}^\alpha\rightarrow 0$ that was used in Section 80.

We extend X_∞ to the maximal backward-time limit and obtain an ancient κ-solution, which contradicts the assumption that the points $(\overline{x}, \overline{t})$ did not have canonical neighborhoods. This proves part (b) of the proposition.

To prove part (c), we can rescale t_0 to 1 and then apply Lemma 70.2; see the end of the proof of Theorem 53.1. The Φ-pinching that we use comes from the Hamilton-Ivey estimate of (B.4). We recall that in the proof of Theorem 53.1 we need to get nonnegative curvature in the region W near the blowup point; this comes from the fact that $\overline{r}^\alpha \to 0$ in the contradiction argument, along with the Hamilton-Ivey pinching. □

This proves the proposition. In what follows, we will want to apply it freely for arbitrary A, provided that t_0 is large enough. To do so, we reduce the function $\overline{\delta}$ used to define the Ricci flow with (r, δ)-cutoff, if necessary, in order to ensure that $\overline{\delta}(t) \leq \overline{\delta}_{2t}(2t)$. Here $\overline{\delta}_{2t}(2t)$ is the quantity $\overline{\delta}_A(2t)$ from Proposition 84.1 evaluated at $A = 2t$.

86. II.6.4. Earlier scalar curvature bounds on smaller balls from lower curvature bounds and volume bounds, in the presence of possible surgeries

In this section we prove Proposition 84.2. We first recall its statement.

Proposition 86.1. *(cf. Proposition II.6.4) There exist $\tau, \overline{r}, C_1 > 0$ and $K < \infty$ with the following property. Suppose that we have a Ricci flow with (r, δ)-cutoff on the time interval $[0, t_0]$, with normalized initial data. Let r_0 satisfy $2C_1 h_{\max}(t_0) \leq r_0 \leq \overline{r}\sqrt{t_0}$, where $h_{\max}(t_0)$ is the maximal cutoff radius for surgeries in $[t_0/2, t_0]$. (If there are no surgeries on $[t_0/2, t_0]$ then we put $h_{\max}(t_0) = 0$.) Assume*
1. The ball $B(x_0, t_0, r_0)$ has sectional curvatures at least $-r_0^{-2}$ at each point.
2. The volume of any subball $B(x, t_0, r) \subset B(x_0, t_0, r_0)$ with any radius $r > 0$ is at least $(1 - \epsilon)$ times the volume of the Euclidean ball of the same radius.

Then the solution is unscathed on $P(x_0, t_0, r_0/4, -\tau r_0^2)$ and satisfies $R < K r_0^{-2}$ there.

Proposition 84.2 is an analog of Theorem 54.2. However, the proof of Proposition 84.2 is more complicated, due to the need to deal with possible surgeries. The idea of the proof is to put oneself in a setting in which one can apply Lemma 82.1. To do this, one needs to first show that the solution is unscathed in a parabolic region $P(x_0, t_0, r_0, -\tau_0 r_0^2)$ and that one has $\mathrm{Rm} \geq -r_0^{-2}$ there.

Proof. The constants C_1, K and τ are fixed numbers, but the requirements on them will be specified during the proof. The number $\overline{r}$ will emerge from the proof, via a contradiction argument.

We first dispose of the case when $r_0 \leq r(t_0)$. Suppose that $r_0 \leq r(t_0)$. We claim that $R \leq r_0^{-2}$ on $B(x_0, t_0, \frac{r_0}{2})$. If not then $R(x, t_0) > r_0^{-2}$ for some $x \in B(x_0, t_0, \frac{r_0}{2})$, and so $R(x, t_0) > r(t_0)^{-2}$. This implies that $R(x, t_0)$ is in a canonical neighborhood, which contradicts the almost-Euclidean-volume assumption on subballs of $B(x_0, t_0, r_0)$.

Thus $R \le r_0^{-2}$ on $B(x_0, t_0, \frac{r_0}{2})$. Lemma 70.1 implies that $R \le 16 r_0^{-2}$ on $P(x_0, t_0, \frac{1}{2} r_0, -\frac{1}{16}\eta^{-1} r_0^2)$. Furthermore, if
(*.1) $C_1 \ \ge\ 100$
then $R \ \le\ \frac{1}{2h^2}$ on $P(x_0, t_0, r_0/4, -\frac{1}{16}\eta^{-1} r_0^2)$. As surgeries only occur when $R \ \ge\ h^{-2}$, there cannot be any surgeries in the region. Hence if we have
(*.2) $K \ \ge\ 200$ and
(*.3) $\tau \le \frac{1}{16}\eta^{-1}$
then there are no counterexamples to the proposition with $r_0 \ \le\ r(t_0)$.

Continuing with the proof of the proposition, suppose that we have a sequence of Ricci flows with (r, δ)-cutoff $\mathcal{M}^\alpha$ satisfying the assumptions of the proposition, with $\overline{r}^\alpha \to 0$, so that the conclusion of the proposition is violated for each $\mathcal{M}^\alpha$. Let t_0^α be the first time when the conclusion is violated for $\mathcal{M}^\alpha$ and let $B(x_0^\alpha, t_0^\alpha, r_0^\alpha)$ be a time-t_0^α ball of smallest radius which provides a counterexample. (Such a ball exists since we have already shown that the proposition holds if $r_0^\alpha \le r(t_0^\alpha)$.) That is, either there is a surgery in $P(x_0^\alpha, t_0^\alpha, r_0^\alpha/4, -\tau (r_0^\alpha)^2)$ or $R \ \ge\ K (r_0^\alpha)^{-2}$ somewhere on $P(x_0^\alpha, t_0^\alpha, r_0^\alpha/4, -\tau (r_0^\alpha)^2)$. From the previous paragraph, $r_0^\alpha \ >\ r(t_0^\alpha)$.

Let $\widehat{\tau}$ be the supremum of the numbers $\widetilde{\tau}$ with the property that for large α, $P(x_0^\alpha, t_0^\alpha, r_0^\alpha, -\widetilde{\tau} (r_0^\alpha)^2)$ is unscathed and $\mathrm{Rm} \ \ge\ -\,(r_0^\alpha)^{-2}$ there.

Lemma 86.2. *$\widehat{\tau}$ is bounded below by the parameter τ_0 of Lemma 82.1, where we take $w = 1 - \epsilon$ in Lemma 82.1.*

Proof. Suppose that $\widehat{\tau} < \tau_0$. Put $\widehat{t^\alpha} \ =\ t_0^\alpha - (1-\epsilon')\,\widehat{\tau}\,(r_0^\alpha)^2$, where ϵ' will eventually be taken to be a small positive number. Applying Lemma 82.1 to the solution on $P(x_0^\alpha, t_0^\alpha, r_0^\alpha, -\,(1-\epsilon')\,\widehat{\tau}\,(r_0^\alpha)^2)$ (see the end of Section 82), the volume of $B(x_0^\alpha, \widehat{t^\alpha}, r_0^\alpha/4)$ is at least $\frac{1}{10}$ of the volume of the Euclidean ball of the same radius. From Lemma 83.1, there is a subball $B(x_1^\alpha, \widehat{t^\alpha}, r^\alpha) \subset B(x_0^\alpha, \widehat{t^\alpha}, r_0^\alpha)$ of radius $r^\alpha \ =\ \theta_0(1/10)\, r_0^\alpha/4$ with the property that all of its subballs have volume at least $(1-\epsilon)$ times the volume of the Euclidean ball of the same radius. The sectional curvature on $B(x_1^\alpha, \widehat{t^\alpha}, r^\alpha)$ is bounded below by $-\,(r^\alpha)^{-2}$. As we can apply the conclusion of the proposition to this subball (in view of its earlier time or smaller radius than $B(x_0^\alpha, t_0^\alpha, r_0^\alpha)$), it follows that the solution on $P(x_1^\alpha, \widehat{t^\alpha}, r^\alpha/4, -\tau (r^\alpha)^2)$ is unscathed and has $R \ <\ K (r^\alpha)^{-2}$. As

$$(86.3) \qquad \frac{(r^\alpha)^2}{\widehat{t^\alpha}} \ \ge\ \frac{(r^\alpha)^2}{(r_0^\alpha)^2}\, \frac{(r_0^\alpha)^2}{t_0^\alpha - \tau_0 (r_0^\alpha)^2} \ =\ \frac{(r^\alpha)^2}{(r_0^\alpha)^2}\, \frac{(r_0^\alpha)^2/t_0^\alpha}{1 - \tau_0 (r_0^\alpha)^2/t_0^\alpha},$$

the fact that $\overline{r}^\alpha \to 0$ as $\alpha \to \infty$ implies that the Hamilton-Ivey pinching improves with α. In particular, for large α, $|\,\mathrm{Rm}\,| < K (r^\alpha)^{-2}$ on $P(x_1^\alpha, \widehat{t^\alpha}, r^\alpha/4, -\tau (r^\alpha)^2)$. Putting $\widetilde{r}_0^\alpha = K^{-1/2} r^\alpha$, if
(*.4) $K^{-1} \ \le\ \frac{1}{10}\,\tau$
then $|\,\mathrm{Rm}\,| \le (\widetilde{r}_0^\alpha)^{-2}$ on the parabolic ball $P(x_1^\alpha, \widetilde{t^\alpha}, \widetilde{r}_0^\alpha, -(\widetilde{r}_0^\alpha)^2)$ for any $\widetilde{t^\alpha} \in [\widehat{t^\alpha} - \frac{1}{2}\tau (r^\alpha)^2, \widehat{t^\alpha}]$. Taking $A \ =\ 100 r_0^\alpha / \widetilde{r}_0^\alpha$, Proposition 84.1(c) now implies that for large α, we have $R \ \le\ K_2(A)\, (\widetilde{r}_0^\alpha)^{-2}$ on $B(x_1^\alpha, \widetilde{t^\alpha}, 100 r_0^\alpha)$. Provided that
(*.5) $K_2(A)\, K\, (\theta_0(1/10))^{-2} \ \le\ \frac{1}{1000}\, C_1^2$
we will have $K_2(A)\, (\widetilde{r}_0^\alpha)^{-2} \ <\ \frac{1}{2}\, h^{-2}$ and so there will not be any surgeries on such balls. Then the length distortion estimates of Lemma 27.8 imply that there is some $c > 0$ so that

$R \le K_2(A)\,(\widetilde{r}_0^\alpha)^{-2}$ on $P(x_0^\alpha, r_0^\alpha, \widehat{t^\alpha}, -c(r_0^\alpha)^2)$. Hamilton-Ivey pinching now implies that for large α, $\mathrm{Rm} \ge -(r_0^\alpha)^{-2}$ on $P(x_0^\alpha, r_0^\alpha, \widehat{t^\alpha}, -c(r_0^\alpha)^2)$. As c can be taken independent of the small number ϵ', taking $\epsilon' \to 0$ we contradict the maximality of $\widehat{\tau}$. □

We can now apply Lemma 82.1 to obtain $R \le K_0 \tau_0^{-1} (r_0^\alpha)^{-2}$ on $P(x_0^\alpha, t_0^\alpha, r_0^\alpha/4, -\tau_0 (r_0^\alpha)^2/2)$. This will give a contradiction provided that
(*.6) $K_0\, \tau_0^{-1} < K/2$,
(*.7) $\tau < \tau_0/2$ and
(*.8) $K_0\, \tau_0^{-1} \le C_1^2$,
where the last condition rules out surgeries in $P(x_0^\alpha, t_0^\alpha, r_0^\alpha/4, -\tau(r_0^\alpha)^2)$.

We choose τ to satisfy ($*$.3) and ($*$.7). We choose K to satisfy ($*$.2), ($*$.4) and ($*$.6). Finally, we choose C_1 to satisfy ($*$.1), ($*$.5) and ($*$.8). This proves the proposition. □

Remark 86.4. In subsequent sections we will want to know that for any $w > 0$, with the notation of Corollary 81.3, we have $\theta^{-1}(w)\, h_{\max}(t_0) \le r(t_0)$ if t_0 is sufficiently large (as a function of w). We can always achieve this by lowering the function $\overline{\delta}(\cdot)$ used to define the Ricci flow with (r, δ)-cutoff so that $\lim_{t_0 \to \infty} \frac{h_{\max}(t_0)}{r(t_0)} = 0$. We will assume hereafter that this is the case.

87. II.7.1. Noncollapsed pointed limits are hyperbolic

In this section we start the analysis of the long-time decomposition into hyperbolic and graph manifold pieces. In the section, $\mathcal{M}$ will denote a Ricci flow with (r, δ)-cutoff whose initial time slice $(\mathcal{M}_0, g(0))$ is compact and has normalized metric.

From Lemma 81.1, if $g(0)$ has positive scalar curvature then the solution goes extinct in a finite time. From Lemma 81.2, these manifolds are understood topologically. If $g(0)$ has nonnegative scalar curvature then either it acquires positive scalar curvature or it is flat, so again the topological type is understood. Hereafter we assume that the flow does not become extinct, and that $R_{\min} < 0$ for all t.

Lemma 87.1. $V(t)\, \left(t + \frac{1}{4}\right)^{-\frac{3}{2}}$ *is nonincreasing in* t.

Proof. Suppose first that the flow is nonsingular. In the case the lemma follows from Lemma 79.11 and the equation

$$\frac{dV}{dt} = -\int_M R\, dV \le -R_{min}\, V. \tag{87.2}$$

If there are surgeries then it only has the effect of causing further decrease in V. □

Definition 87.3. Put $\overline{V} = \lim_{t \to \infty} V(t)\, \left(t + \frac{1}{4}\right)^{-\frac{3}{2}}$ and $\hat{R}(t) = R_{min}(t)\, V(t)^{\frac{2}{3}}$.

Lemma 87.4. *On any time interval which is free of singular times, and on which* $R_{min}(t) \le 0$ *for all* t *(which we are assuming), we have*

$$\frac{d\hat{R}}{dt} \ge \frac{2}{3}\, \hat{R}\, V^{-1} \int_M (R_{min} - R)\, dV. \tag{87.5}$$

Proof. From (B.1), $\frac{dR_{min}}{dt} \geq \frac{2}{3} R_{min}^2$. Then

$$\frac{d\hat{R}}{dt} = \frac{dR_{min}}{dt} V^{\frac{2}{3}} + \frac{2}{3} R_{min} V^{-\frac{1}{3}} \frac{dV}{dt} \geq \frac{2}{3} R_{min}^2 V^{\frac{2}{3}} - \frac{2}{3} R_{min} V^{-\frac{1}{3}} \int_M R\, dV, \tag{87.6}$$

from which the lemma follows. □

Corollary 87.7. *If $R_{min}(t) \leq 0$ for all t (which we are assuming) then $\hat{R}(t)$ is nondecreasing.*

Proof. If $\mathcal{M}$ is a nonsingular flow then the corollary follows from Lemma 87.4. If there are surgeries then it only has the effect of decreasing $V(t)$, and so possibly increasing $\hat{R}(t)$ (since $R_{min}(t) \leq 0$). □

Put $\overline{R} = \lim_{t\to\infty} \hat{R}(t)$.

Lemma 87.8. *If $\overline{V} > 0$ then $\overline{R}\,\overline{V}^{-2/3} = -\frac{3}{2}$.*

Proof. Suppose that $\overline{V} > 0$. Using Lemma 79.11,
(87.9)

$$\overline{R}\,\overline{V}^{-2/3} = \lim_{t\to\infty} R_{min}(t) V(t)^{\frac{2}{3}} \left(V(t) \left(t + \frac{1}{4} \right)^{-\frac{3}{2}} \right)^{-\frac{2}{3}} = \lim_{t\to\infty} \left(t + \frac{1}{4} \right) R_{min}(t) \geq -\frac{3}{2}.$$

In particular, there is a limit as $t \to \infty$ of $\left(t + \frac{1}{4}\right) R_{min}(t)$. Suppose that

$$\overline{R}\,\overline{V}^{-2/3} = \lim_{t\to\infty} \left(t + \frac{1}{4} \right) R_{min}(t) = c > -\frac{3}{2}. \tag{87.10}$$

Combining this with (87.2) gives that for any $\mu > 0$, $V(t) \leq$ const. $t^{\mu - c}$ whenever t is sufficiently large. Then $V(t) t^{-3/2} \leq$ const. $t^{-(c+3/2-\mu)}$. Taking $\mu = \frac{1}{2}(c + \frac{3}{2})$, we contradict the assumption that $\overline{V} > 0$. □

From the proof of Lemma 87.8, if $\overline{V} > 0$ then $R_{min}(t) \sim -\frac{3}{2t}$.

The next proposition shows that a long-time limit will necessarily be hyperbolic.

Proposition 87.11. *Given the flow $\mathcal{M}$, suppose that we have a sequence of parabolic neighborhoods $P(x^\alpha, t^\alpha, r\sqrt{t^\alpha}, -r^2 t^\alpha)$, for $t^\alpha \to \infty$ and some fixed $r \in (0,1)$, such that the scalings of the parabolic neighborhoods with factor t^α smoothly converge to some limit solution $(\mathcal{M}_\infty, (\overline{x}, 1), g_\infty(\cdot))$ defined in a parabolic neighborhood $P(\overline{x}, 1, r, -r^2)$. Then $g_\infty(t)$ has constant sectional curvature $-\frac{1}{4t}$.*

Proof. Suppose first that the flow is surgery-free. Because of the assumed existence of the limit $(\mathcal{M}_\infty, (\overline{x}, 1), g_\infty(\cdot))$, the original solution $\mathcal{M}$ has $\overline{V} > 0$. We claim that the scalar curvature on $P(\overline{x}, 1, r, -r^2)$ is spatially constant. If not then there are numbers $c < 0$ and $s_0, \mu > 0$ so that

$$\int_{B(\overline{x},s,r)} (R'_{min}(s) - R(x,s))\, dV \leq c \tag{87.12}$$

whenever $s \in (s_0 - \mu, s_0 + \mu) \subset [1 - r^2, 1]$, where $R'_{min}(s)$ is the minimum of R over $B(\overline{x}, s, r)$. Then for large α,

$$\int_{B(x^\alpha, st^\alpha, r\sqrt{t^\alpha})} (R'_{min}(st^\alpha) \; - \; R(x, st^\alpha)) \, dV \; < \; \frac{c}{2} \sqrt{t^\alpha}, \tag{87.13}$$

where $R'_{min}(st^\alpha)$ is now the minimum of R over $B(x^\alpha, st^\alpha, r\sqrt{t^\alpha})$. Thus

$$\int_{B(x^\alpha, st^\alpha, r\sqrt{t^\alpha})} (R_{min}(st^\alpha) \; - \; R(x, st^\alpha)) \, dV \; < \; \frac{c}{2} \sqrt{t^\alpha}. \tag{87.14}$$

After passing to a subsequence, we can assume that $\frac{t^{\alpha+1}}{t^\alpha} > \frac{s_0+\mu}{s_0-\mu}$ for all α. From (87.5),

(87.15)

$$\begin{aligned} \overline{R} \; - \; \hat{R}(0) \; &\geq \; \frac{2}{3} \int_0^\infty \hat{R}(t) \, V(t)^{-1} \int_M (R_{min}(t) - R(x,t)) \, dV(x) \, dt \\ &\geq \; \frac{2}{3} \sum_\alpha t^\alpha \int_{s_0-\mu}^{s_0+\mu} \hat{R}(st^\alpha) \, V(st^\alpha)^{-1} \int_M (R_{min}(st^\alpha) - R(x, st^\alpha)) \, dV(x) \, ds \\ &\geq \; \frac{2}{3} \sum_\alpha t^\alpha \int_{s_0-\mu}^{s_0+\mu} \hat{R}(st^\alpha) \, V(st^\alpha)^{-1} \int_{B(x^\alpha, st^\alpha, r\sqrt{t^\alpha})} (R_{min}(st^\alpha) - R(x, st^\alpha)) \, dV(x) \, ds. \end{aligned}$$

Using the definitions of $\overline{V}$ and $\overline{R} = -\frac{3}{2}\overline{V}^{\frac{2}{3}}$, along with (87.14), it follows that the right-hand side of (87.15) is infinite. This contradicts the fact that $\overline{R} < \infty$.

Thus R is spatially constant on $P(\overline{x}, 1, r, -r^2)$. As $R_{min}(t) \sim -\frac{3}{2t}$ on M, we know that the scalar curvature R at $(x,t) \in P(\overline{x}, 1, r, -r^2)$, which only depends on t, satisfies $R(t) \geq -\frac{3}{2t}$. It does not immediately follow that the scalar curvature on $P(\overline{x}, 1, r, -r^2)$ equals $-\frac{3}{2t}$, as $R_{min}(t)$ is the minimum of the scalar curvature on all of M. However, if the scalar curvature is not identically $-\frac{3}{2t}$ on $P(\overline{x}, 1, r, -r^2)$ then again we can find $c < 0$ and $s_0, \mu > 0$ so that for large α, (87.14) holds for $s \in (s_0 - \mu, s_0 + \mu) \subset [1 - r^2, 1]$. Again we get a contradiction using (87.5). Thus $R(t) = -\frac{3}{2t}$ on $P(\overline{x}, 1, r, -r^2)$. Then from (B.1), each time-slice of $P(\overline{x}, 1, r, -r^2)$ has an Einstein metric. Thus the sectional curvature on $P(\overline{x}, 1, r, -r^2)$ is $-\frac{1}{4t}$.

The argument goes through if one allows surgeries. The main ingredient was the monotonicity formulas, which still hold if there are surgeries. Note that for large α there are no surgeries in $P(x^\alpha, t^\alpha, r\sqrt{t^\alpha}, -r^2 t^\alpha)$ by assumption. □

88. II.7.2. Noncollapsed regions with a lower curvature bound are almost hyperbolic on a large scale

In this section it is shown that for fixed $A, r, w > 0$ and large time t_0, if $B(x_0, t_0, r\sqrt{t_0}) \subset \mathcal{M}^+_{t_0}$ has volume at least $wr^3 t_0^{\frac{3}{2}}$ and sectional curvatures at least $-r^{-2}t_0^{-1}$ then the Ricci flow on the parabolic neighborhood $P(x_0, t_0, Ar\sqrt{t_0}, Ar^2 t_0)$ is close to the flow on a hyperbolic manifold.

We retain the assumptions of the previous section.

Lemma 88.1. *(cf. Lemma II.7.2)*
(a) Given $w, r, \xi > 0$ one can find $T = T(w, r, \xi) > \infty$ such that if the ball $B(x_0, t_0, r\sqrt{t_0}) \subset \mathcal{M}^+_{t_0}$ at some time $t_0 \geq T$ has volume at least $wr^3 t_0^{\frac{3}{2}}$ and sectional curvatures at least $-r^{-2}t_0^{-1}$ then the curvature at (x_0, t_0) satisfies

$$|2tR_{ij}(x_0, t_0) + g_{ij}|^2 \ = \ (2tR_{ij}(x_0, t_0) + g_{ij})(2tR^{ij}(x_0, t_0) + g^{ij}) \ < \ \xi^2. \tag{88.2}$$

(b) Given in addition $A < \infty$ and allowing T to depend on A, we can ensure (88.2) for all points in $B(x_0, t_0, Ar\sqrt{t_0})$.
(c) The same is true for $P(x_0, t_0, Ar\sqrt{t_0}, Ar^2 t_0)$.

Note that the time T will depend on the initial metric.

Proof. To prove (a), suppose that there is a sequence of points (x_0^α, t_0^α) with $t_0^\alpha \to \infty$ that provide a counterexample. We wish to apply Corollary 81.3 with the parameter r_0 of the corollary equal to $r\sqrt{t_0^\alpha}$. Putting

$$\widehat{w} \ = \ \left(\min_{x \in [0,1]} \frac{\int_0^x \sinh^2(s)\, ds}{x^3 \int_0^1 \sinh^2(s)\, ds} \right) w, \tag{88.3}$$

if $r > \overline{r}(\widehat{w})$, where $\overline{r}$ is from Corollary 81.3, then the hypotheses of the lemma will still be satisfied upon replacing w by $\widehat{w}$ and r by $\overline{r}(\widehat{w})$. Thus after redefining w, if necessary, we may assume that $r \leq \overline{r}(w)$. As the function $h_{\max}(t)$ is nonincreasing, if t_0^α is sufficiently large then $\theta^{-1}(w) h_{\max}(t_0^\alpha) \ \leq \ r\sqrt{t_0^\alpha}$. Using Corollary 81.3 with a redefinition of w, we can take a convergent pointed subsequence as $\alpha \to \infty$ of the t_0^α-rescalings, whose limit is defined in an abstract parabolic neighborhood. From Proposition 87.11 the limit will be hyperbolic, which is a contradiction.

For part (b), Corollary 81.3 gives a bound $R \leq K r_0^{-2}$ in the unscathed parabolic neighborhood $P(x_0, t_0, r_0/4, -\tau r_0^2)$, where $r_0 = \min(r, \overline{r}(w'))\sqrt{t_0}$. We apply Proposition 84.1 to the parabolic neighborhood $P(x_0, t_0, r_0', -(r_0')^2)$ where $K r_0^{-2} \ = \ (r_0')^{-2}$. By Proposition 84.1(b), each point $y \in B(x_0, t_0, Ar\sqrt{t_0})$ with scalar curvature at least $Q = K'(A) r_0^{-2}$ has a canonical neighborhood. Suppose that there is such a point. From part (a) we have $R(x_0, t_0) < 0$, so along a geodesic from x_0 to y there will be some point $x_0' \in B(x_0, t_0, Ar\sqrt{t_0})$ with scalar curvature Q. It also has a canonical neighborhood, necessarily of type (a) or (b). We can apply part (a) to a ball around x_0' with a radius on the order of $(K'(A))^{-1/2} r_0$, and with a value of w coming from the canonical neighborhood condition, to get a contradiction for large t_0. (Note that $(K'(A))^{-1/2} r_0$ is proportionate to $\sqrt{t_0}$.) Thus $R \ \leq \ K'(A) r_0^{-2}$ on $B(x_0, t_0, Ar\sqrt{t_0})$. If T is large enough then the Φ-almost nonnegative curvature implies that $|\operatorname{Rm}| \ \leq \ K'(A) r_0^{-2}$. Then the noncollapsing in Proposition 84.1(a) gives a lower local volume bound. Hence we can apply part (a) of the lemma to appropriate-sized balls in $B(x_0, t_0, \frac{A}{2} r\sqrt{t_0})$. As A is arbitrary, this proves (b) of the lemma.

For part (c), without loss of generality we can take ξ small. Suppose that the claim is not true. Then there is a point (x_0, t_0) that satisfies the hypotheses of the lemma but for which there is a point $(x_1, t_1) \in P(x_0, t_0, Ar\sqrt{t_0}, Ar^2 t_0)$ with $|2t_1 R_{ij}(x_1, t_1) + g_{ij}| \geq \xi$. Without loss of generality, we can take (x_1, t_1) to be a first such point in $P(x_0, t_0, Ar\sqrt{t_0}, Ar^2 t_0)$. By part (b), $t_1 > t_0$. Then $|2tR_{ij} + g_{ij}| \leq \xi$ on $P(x_0, t_0, Ar\sqrt{t_0}, t_1 - t_0)$. If ξ is small then this region has negative sectional curvature and there are no surgeries in the region.

Using the length distortion estimates of Section 27, we can find $r' = r'(r, A) > 0$ so that the sectional curvature is bounded below by $-\,(r')^{-2}\,t_1^{-1}$ on $B(x_0, t_1, r'\sqrt{t_1})$. Also, by the evolution of volume under Ricci flow, there will be a $w' = w'(r, w, \xi, A)$ so that the volume of $B(x_0, t_1, r'\sqrt{t_1})$ is bounded below by $w'(r')^3(t_1)^{\frac{3}{2}}$. Thus for large t_0 we can apply (b) to $B(x_0, t_1, A'r'\sqrt{t_1})$ with an appropriate choice of A' to obtain a contradiction. □

89. II.7.3. Thick-thin decomposition

This section is concerned with the large-time decomposition of the manifold in "thick" and "thin" parts.

Definition 89.1. For $x \in \mathcal{M}_t^+$, let $\rho(x, t)$ be the unique number $\rho \in (0, \infty)$ such that $\inf_{B^+(x,t,\rho)} \mathrm{Rm} = -\,\rho^{-2}$, if such a ρ exists, and put $\rho(x, t) = \infty$ otherwise.

The function $\rho(x, t)$ is well-defined because $\mathcal{M}_t^+$ is a compact smooth Riemannian manifold, so for fixed $(x, t) \in \mathcal{M}_t^+$ the quantity $\inf_{B^+(x,t,\rho)} \mathrm{Rm}$ is a continuous nonincreasing function of ρ which is negative for sufficiently large ρ if and only if (x, t) lies in a connected component with negative sectional curvature somewhere; on the other hand the function $-\rho^{-2}$ is continuous and strictly increasing. We note that when it is finite, the quantity $\rho(x, t)$ may be larger than the diameter of the component of $\mathcal{M}_t$ containing (x, t).

As an example, if $\mathcal{M}$ is the flow on a manifold M with spatially constant negative curvature then for large t, $\rho(x, t) \sim 2\sqrt{t}$ uniformly on M. The "thin" part of $\mathcal{M}_t$, in the sense of hyperbolic geometry, can then be characterized as the points x so that $\mathrm{vol}(B(x, t, \rho(x, t)) < w\,\rho^3(x, t)$, for an appropriate constant w.

Lemma 89.2. *For any $w > 0$ we can find $\overline{\rho} = \overline{\rho}(w) > 0$ and $\overline{T} = \overline{T}(w)$ such that if $t \geq \overline{T}$ and $\rho(x, t) < \overline{\rho}\sqrt{t}$ then*

$$\mathrm{vol}(B(x, t, \rho(x, t))) \;<\; w\rho^3(x, t). \tag{89.3}$$

Proof. If the lemma is not true then there is a sequence (x^α, t^α) with $t^\alpha \to \infty$, $\rho(x^\alpha, t^\alpha)(t^\alpha)^{-1/2} \to 0$ and $\mathrm{vol}(B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))) \geq w\rho^3(x^\alpha, t^\alpha)$. The first step is to apply Corollary 81.3, but we need to know that for large α we have $\rho(x^\alpha, t^\alpha) \;\geq\; \theta^{-1}(w)\,h_{\max}(t^\alpha)$, where $\theta(w)$ and $h_{\max}$ are from Corollary 81.3. Suppose that this is not the case. Then after passing to a subsequence we have $\rho(x^\alpha, t^\alpha) \;<\; \theta^{-1}(w)\,h_{\max}(t^\alpha) \leq r(t^\alpha)$ for all α, where we used Remark 86.4 in the last inequality. There are points $x^{\alpha,\prime} \in \overline{B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))}$ with a sectional curvature equal to $-\,\rho^{-2}(x^\alpha, t^\alpha)$. Applying the Hamilton-Ivey pinching estimate of (B.4) with $X^\alpha = \rho^{-2}(x^\alpha, t^\alpha)$, and using the fact that $\lim_{\alpha\to\infty} t^\alpha X^\alpha \;=\; \infty$, gives

$$\lim_{\alpha\to\infty} R(x^{\alpha,\prime}, t^\alpha)\,\rho^2(x^\alpha, t^\alpha) \;=\; \infty. \tag{89.4}$$

We claim that the curvatures at the centers of the balls satisfy

$$\lim_{\alpha\to\infty} R(x^\alpha, t^\alpha)\,\rho^2(x^\alpha, t^\alpha) \;=\; \infty. \tag{89.5}$$

Suppose not. Then there is some number $C \in (0, \infty)$ so that after passing to a subsequence, $R(x^\alpha, t^\alpha)\,\rho^2(x^\alpha, t^\alpha) \;\leq\; C$ for all α. By continuity and (89.4), after passing to another subsequence we can assume that there is a point $x^{\alpha\prime\prime}$ on a time-t^α geodesic segment between

x^α and $x^{\alpha\prime}$ so that $R(x^{\alpha\prime\prime}, t^\alpha)\, \rho^2(x^\alpha, t^\alpha) = 2C$, for all α. We now apply Lemma 70.2 around $(x^{\alpha\prime\prime}, t^\alpha)$ to get a contradiction to (89.4). (More precisely, we apply a version of Lemma 70.2 that applies along geodesics, as in Claim 2 of II.4.2.) In applying Lemma 70.2, we use the fact that $\lim_{\alpha\to\infty} t^\alpha\, \rho^{-2}(x^\alpha, t^\alpha) = \infty$ in order to say that $\lim_{\alpha\to\infty} \frac{\Phi(2C\rho^{-2}(x^\alpha,t^\alpha))}{2C\rho^{-2}(x^\alpha,t^\alpha)} = 0$, with the notation of Lemma 70.2.

Making a similar argument centered at other points in $B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$, we deduce that

$$\lim_{\alpha\to\infty} \left(\inf R\big|_{B(x^\alpha,t^\alpha,\rho(x^\alpha,t^\alpha))} \right) \rho^2(x^\alpha, t^\alpha) = \infty. \tag{89.6}$$

In particular, since $\rho(x^\alpha, t^\alpha) \leq r(t^\alpha)$, if α is large then each point $y^\alpha \in B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$ is the center of a canonical neighborhood. As $\alpha \to \infty$, the intrinsic scales $R(y^\alpha, t^\alpha)^{-1/2}$ become arbitrarily small compared to $\rho(x^\alpha, t^\alpha)$. However, by Lemma 83.1 there is a subball $B^{\alpha,\prime}$ of $B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$ with radius $\theta_0(w)\rho(x^\alpha, t^\alpha)$ so that every subball of $B^{\alpha,\prime}$ has almost Euclidean volume. This contradicts the existence of a small canonical neighborhood around each point of $B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$.

We now know that for large α, $\rho(x^\alpha, t^\alpha) \geq \theta^{-1}(w)\, h_{\max}(t^\alpha)$. Thus we can apply Corollary 81.3 to get an unscathed solution on the parabolic neighborhood $P(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha)/4, -\tau\rho^2(x^\alpha, t^\alpha))$, with $R < K_0\, \rho(x^\alpha, t^\alpha)^{-2}$ there. Applying Proposition 84.1(c), along with the fact that $\lim_{\alpha\to\infty} t^\alpha \rho^{-2}(x^\alpha, t^\alpha) = \infty$, gives an estimate $R \leq K_2 \rho(x^\alpha, t^\alpha)^{-2}$ on $B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$. But then for large α, the Hamilton-Ivey pinching gives $\mathrm{Rm} > -\frac{1}{2}\rho(x^\alpha, t^\alpha)^{-2}$ on $B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))$, which is a contradiction. □

Remark 89.7. Another approach to the above proof would be to use the canonical neighborhood at the center (x^α, t^α) of the ball, along with the Bishop-Gromov inequality, to contradict the fact that $\mathrm{vol}(B(x^\alpha, t^\alpha, \rho(x^\alpha, t^\alpha))) \geq w\rho^3(x^\alpha, t^\alpha)$. For this to work we would have to know that the relative volume $(\epsilon^2\, R(x^\alpha, t^\alpha))^{\frac{3}{2}}\ \mathrm{vol}(B(x^\alpha, t^\alpha, \epsilon^{-1} R(x^\alpha, t^\alpha)^{-\frac{1}{2}})$ of the canonical neighborhood (of type (a) or (b)) around (x^α, t^α) is small compared to w. This will be the case if we take the constant ϵ to be small enough, but in a w-dependent way. Although ϵ is supposed to be a universal constant, this approach will work because when characterizing the graph manifold part in Section 92, w can be taken to be a small but fixed constant.

Definition 89.8. The *w-thin part* $M^-(w,t) \subset \mathcal{M}_t^+$ is the set of points $x \in M$ so that either $\rho(x,t) = \infty$ or

$$\mathrm{vol}(B(x, t, \rho(x,t))) < w\, (\rho(x,t))^3. \tag{89.9}$$

The *w-thick part* is $M^+(w,t) = \mathcal{M}_t^+ - M^-(w,t)$.

Lemma 89.10. *Given $w > 0$, there are $w' = w'(w) > 0$ and $T' = T'(w) < \infty$ so that taking $r = \overline{\rho}(w)$ (with reference to Lemma 89.2), if $x_0 \in M^+(w,t)$ and $t_0 \geq T'$ then $B(x_0, t_0, r\sqrt{t_0})$ has volume at least $w' r^3 t_0^{\frac{3}{2}}$ and sectional curvature at least $-\, r^{-2}\, t_0^{-1}$.*

Proof. Suppose that $x_0 \in M^+(w, t_0)$. From Lemma 89.2, if t_0 is big enough (as a function of w) then $\rho(x_0, t_0) \geq r\,\sqrt{t_0}$. As $\mathrm{Rm} \geq -\,\rho(x_0, t_0)^{-2}$ on $B(x_0, t_0, \rho(x_0, t_0))$, we have $\mathrm{Rm} \geq -\, r^{-2}\, t_0^{-1}$ on $B(x_0, t_0, r\sqrt{t_0})$. As $\mathrm{vol}(B(x_0, t_0, \rho(x_0, t_0))) \geq w\, (\rho(x_0, t_0))^3$, the Bishop-Gromov inequality gives a lower bound on $\left(r\sqrt{t_0}\right)^{-3}\ \mathrm{vol}(B(x_0, t_0, r\sqrt{t_0}))$ in terms of w. □

90. Hyperbolic rigidity and stabilization of the thick part

Lemma 89.10 implies that Lemma 88.1 applies to $M^+(w,t)$ if t is sufficiently large (as a function of w). That is, if one takes a sequence of points in the w-thick parts at a sequence of times tending to infinity then the pointed time slices subconverge, modulo rescaling the metrics by t^{-1}, to complete finite volume hyperbolic manifolds with sectional curvatures equal to $-\frac{1}{4}$. (The $-\frac{1}{4}$ comes from the Ricci flow equation along with the equation $g(t) = t\, g(1)$ for the rescaled limit, which implies that $g(1)$ has Einstein constant $-\frac{1}{2}$.) In what follows we will take the word "hyperbolic" for a 3-manifold to mean "constant sectional curvature $-\frac{1}{4}$". The next step, following Hamilton [34], is to show that for large time the picture stabilizes, i.e. the limits are unique in a strong sense.

Proposition 90.1. *There exist a number $T_0 < \infty$, a nonincreasing function $\alpha : [T_0, \infty) \to (0,\infty)$ with $\lim_{t\to\infty} \alpha(t) = 0$, a (possibly empty) collection $\{(H_1,x_1),\ldots,(H_k,x_k)\}$ of complete connected pointed finite-volume hyperbolic 3-manifolds and a family of smooth maps*

$$f(t) : B_t = \bigcup_{i=1}^{k} B\left(x_i, \frac{1}{\alpha(t)}\right) \longrightarrow \mathcal{M}_t, \tag{90.2}$$

defined for $t \in [T_0,\infty)$, such that

1. $f(t)$ is close to an isometry:

$$\|t^{-1} f(t)^* g_{\mathcal{M}_t} - g_{B_t}\|_{C^{\frac{1}{\alpha(t)}}} < \alpha(t), \tag{90.3}$$

2. $f(t)$ defines a smooth family of maps which changes slowly with time:

$$|\dot{f}(p,t)| < \alpha(t) t^{-\frac{1}{2}} \tag{90.4}$$

for all $p \in B_t$, where $\dot{f}$ refers to the time derivative (as defined with admissible curves), and

3. $f(t)$ parametrizes more and more of the thick part: $M^+(\alpha(t),t) \subset \operatorname{im}(f(t))$ for all $t \geq T_0$.

Remark 90.5. The analogous statement in [52, Section 7.3] is in terms of a fixed w. That is, for a given w one considers pointed limits of $\{(\mathcal{M}^+_{t_j}, (x_j,t_j), t_j^{-1} g(t_j))\}_{j=1}^\infty$ with $\lim_{j\to\infty} t_j = \infty$ and the basepoint satisfying $x_j \in M^+(w,t_j) \subset \mathcal{M}^+_{t_j}$ for all j. Considering the possible limit spaces in a certain order, as described below, one extracts complete pointed finite-volume hyperbolic manifolds $\{(H_i,x_i)\}_{i=1}^k$ with x_i in the w-thick part of H_i. There is a number $w_0 > 0$ so that as long as $w \leq w_0$, the hyperbolic manifolds H_i are independent of w. For any $w' > 0$, as time goes on the w'-thick part of $\bigcup_{i=1}^k H_i$ better approximates $M^+(w',t)$. Hence the formulation of Proposition 90.1 is equivalent to that of [52, Section 7.3].

Rather than proving Proposition 90.1 using harmonic maps as in [34], we give a simple proof using smooth compactness and a smoothing argument. Roughly speaking, the idea is to exploit a variant of Mostow rigidity to show that for large t, the components of the w-thick part change slowly with time, and are close to hyperbolic manifolds which are isolated (due to a refinement of Mostow-Prasad rigidity). This forces them to eventually stabilize.

Definition 90.6. If (X,x) and (Y,y) are pointed smooth Riemannian manifolds and $\epsilon > 0$ then (X,x) is ϵ-close to (Y,y) if there is a pointed map $f : (X,x) \to (Y,y)$ such that

$$f\Big|_{\overline{B(x,\epsilon^{-1})}} : \overline{B(x,\epsilon^{-1})} \to Y \tag{90.7}$$

is a diffeomorphism onto its image and

$$\| f^* g_Y - g_X \|_{C^{\epsilon^{-1}}} < \epsilon, \tag{90.8}$$

where the norm is taken on $\overline{B(x,\epsilon^{-1})}$. Note that nothing is required of f on the complement of $\overline{B(x,\epsilon^{-1})}$. Such a map f is called an *ϵ-approximation.*

We will sometimes refer to a partially defined map $f : (X,x) \supset (W,x) \to (Y,y)$ as an ϵ-approximation provided that W contains $\overline{B(x,\epsilon^{-1})}$ and the conditions above are satisfied. By convention we will permit X and Y to be disconnnected, in which case ϵ-closeness only says something about the components containing the basepoints. We say that two maps $f_1, f_2 : (X,x) \to Y$ (not necessarily basepoint-preserving) are ϵ-close if

$$\sup_{p \in \overline{B(x,\epsilon^{-1})}} d_Y(f_1(p), f_2(p)) < \epsilon. \tag{90.9}$$

We recall some facts about hyperbolic manifolds. There is a constant $\mu_0 > 0$, the Margulis constant, such that if X is a complete connected finite-volume hyperbolic 3-manifold (orientable, as usual), $\mu \leq \mu_0$, and

$$X_\mu = \{ x \in X \mid \mathrm{InjRad}(X,x) \geq \mu \} \tag{90.10}$$

is the μ-thick part of X, then X_μ is a nonempty compact manifold-with-boundary whose complement U is a finite union of components $U_1, \ldots, U_k$, where each U_i is isometric either to a geodesic tube around a closed geodesic or to a cusp. In particular, X_μ is connected and there is a one-to-one correspondence between the boundary components of X_μ and the "thin" components U_i. For each i, let $\rho_i : \overline{U_i} \to \mathbb{R}$ denote either the distance function from the core geodesic or a Busemann function, in the tube and cusp cases respectively. In the latter case we normalize ρ_i so that $\rho_i^{-1}(0) = \partial U_i$. (The Busemann function goes to $-\infty$ as one goes down the cusp.) The *radial direction* is the direction field on $U_i - \mathrm{core}(U_i)$ defined by $\nabla \rho_i$, where $\mathrm{core}(U_i)$ is the core geodesic when U_i is a geodesic tube and the empty set otherwise.

Lemma 90.11. *Let (X,x) be a pointed complete connected finite-volume hyperbolic 3-manifold. Then for each $\zeta > 0$ there exists $\xi > 0$ such that if X' is a complete finite-volume hyperbolic manifold with at least as many cusps as X, and $f : (X,x) \to X'$ is a ξ-approximation, then there is an isometry $\hat{f} : (X,x) \to X'$ which is ζ-close to f.*

This was stated as Theorem 8.1 in [34] as going back to the work of Mostow. We give a proof here. The hypothesis about cusps is essential because every pointed noncompact finite-volume hyperbolic 3-manifold (X,x) is a pointed limit of a sequence $\{(X_i,x_i)\}_{i=1}^\infty$ of compact hyperbolic manifolds. Hence for every $\xi > 0$, if i is sufficiently large then there is a ξ-approximation $f : (X,x) \to (X_i,x_i)$, but there is no isometry from X to X_i.

Proof. The main step is to show that for the fixed (X, x), if ξ is sufficiently small then for any ξ-approximation $f : (X, x) \to X'$ satisfying the hypotheses of the lemma, the manifolds X and X' are diffeomorphic. The proof of this will use the Margulis thick-thin decomposition. The rest of the assertion then follows readily from Mostow-Prasad rigidity [47, 54].

Pick $\mu_1 \in (0, \mu_0)$ so that $X - X_{\mu_1}$ consists only of cusps $U_1, \ldots, U_k$. The thick part X_{μ_1} is compact and connected. Given $\xi > 0$, let $f : (B(x, \xi^{-1}), x) \to (X', x')$ be a ξ-approximation as in (90.8). The intuitive idea is that because of the compactness of X_{μ_1}, if ξ is sufficiently small then $f\big|_{X_{\mu_1}}$ is close to being an isometry from X_{μ_1} to its image. Then $f(X_{\mu_1})$ is close to a connected component of the thick part X'_{μ_1} of X'. As $f(X_{\mu_1})$ and X'_{μ_1} are connected, this means that $f(X_{\mu_1})$ is close to X'_{μ_1}. We will show that in fact X_{μ_1} is diffeomorphic to X'_{μ_1}. The boundary components of X_{μ_1} correspond to the cusps of X and the boundary components of X'_{μ_1} correspond to the connected components of $X' - X'_{\mu_1}$. As X' has at least as many cusps as X by assumption, it follows that the connected components of $X' - X'_{\mu_1}$ are all cusps. Hence X and X' are diffeomorphic.

In order to show that X_{μ_1} is diffeomorphic to X'_{μ_1}, we will take a larger region $W \supset X_{\mu_1}$ that is diffeomorphic to X_{μ_1} and show that $f(W)$ can be isotoped to X'_{μ_1} by sliding it inward along the radial direction. More precisely, in each cusp U_i put $V_i = \rho_i^{-1}([-3L, -L])$, where $L \gg 1$ is large enough that every cuspidal torus $\rho_i^{-1}(s)$ with $s \in [-3L, -L]$ has diameter much less than one. Let $W \subset X$ be the complement of the open horoballs at height $-2L$, i.e.

$$W = X - \bigcup_{i=1}^{k} \rho_i^{-1}(-\infty, -2L). \tag{90.12}$$

When ξ is sufficiently small, f will preserve injectivity radius to within a factor close to 1 for points $p \in B(x, \xi^{-1})$ with $d(p, \partial B(x, \xi^{-1})) > 2 \operatorname{InjRad}(X, p)$. Therefore when ξ is small, f will map each V_i into $X' - X'_{\mu_1}$, and hence into one of the connected components U'_{k_i} of $X' - X'_{\mu_1}$. Let Z'_i be the image of $Z_i = \rho_i^{-1}(-2L)$ under f. Note that $d(\mathrm{core}(U'_{k_i}), Z'_i) \gtrsim L$ (if $\mathrm{core}(U'_{k_i}) \neq \emptyset$), for otherwise $f^{-1}(\mathrm{core}(U'_{k_i}))$ would be a closed curve with small diameter and curvature lying in U_i, which contradicts the fact that the horospheres have principal curvatures $-\frac{1}{2}$ (because of our normalization that the sectional curvatures are $-\frac{1}{4}$). Thus for each point $z' \in Z'_i$ there is a minimizing radial geodesic segment γ' passing through z' with $d(z', \partial \gamma') \gtrsim L$. The preimage of γ' under f is a curve $\gamma \subset X$ with small curvature and length $\gtrsim L$ passing through Z_i. This forces the direction of γ to be nearly radial and so transverse to Z_i. Hence Z'_i is transverse to the radial direction in U'_{k_i}. Combining this with the fact that Z'_i is embedded implies that Z'_i is isotopic in $\overline{U'_{k_i}}$ to $\partial U'_{k_i}$. It follows that $f(W)$ is isotopic to X'_{μ_1}. Then by the preceding argument involving counting the number of cusps, X and X' are diffeomorphic. We apply Mostow-Prasad rigidity [47, 54] to deduce that X is isometric to X'.

We now claim that for any $\zeta > 0$, if ξ is sufficiently small then the map f is ζ-close to an isometry from X to X'. Suppose not. Then there are a number $\zeta > 0$ and a sequence of $\frac{1}{i}$-approximations $f_i : (X, x) \to (X'_i, x'_i)$ so that none of the f_i's are ζ-close to any isometry from (X, x) to (X'_i, x'_i). Taking a convergent subsequence of the maps f_i gives a

limit isometry $f_\infty : (X,x) \to (X'_\infty, x'_\infty)$. From what has already been proven, for large i we know that X'_i is isometric to X, and so isometric to X'_∞. This is a contradiction. □

Recall the statement of Lemma 88.1.

Definition 90.13. Given $w > 0$, let Λ_w be the space of complete pointed finite-volume hyperbolic 3-manifolds that arise as pointed limits of sequences $\{(\mathcal{M}^+_{t_i}, (x_i,t_i), t_i^{-1} g(t_i))\}_{i=1}^\infty$ with $\lim_{i\to\infty} t_i = \infty$ and the basepoint (x_i,t_i) satisfying $(x_i,t_i) \in M^+(w,t_i) \subset \mathcal{M}^+_{t_i}$ for all i.

The space Λ_w is compact in the smooth pointed topology. Any element of Λ_w has volume at most $\overline{V}$, the latter being defined in Definition 87.3.

The next lemma summarizes the content of Lemma 88.1.

Lemma 90.14. *Given $w > 0$, there is a decreasing function $\beta : [0,\infty) \to (0,\infty]$ with $\lim_{s\to\infty} \beta(s) = 0$ such that if $(x,t) \in M^+(w,t) \subset \mathcal{M}^+_t$, and Z_t denotes the forward time slice $\mathcal{M}^+_t$ rescaled by t^{-1}, then*

1. Some $(X,x) \in \Lambda_w$ is $\beta(t)$-close to $(Z_t,(x,t))$.

2. $B(x,t,\beta(t)^{-1}\sqrt{t}) \subset \mathcal{M}^+_t$ is unscathed on the interval $[t,2t]$ and if $\gamma : [t,2t] \to \mathcal{M}$ is a static curve starting at (x,t), $\bar{t} \in [t,2t]$, then the map

$$B(x,t,\beta(t)^{-1}\sqrt{t}) \to P(x,t,\beta(t)^{-1}\sqrt{t},t) \cap \mathcal{M}_{\bar{t}} \tag{90.15}$$

defined by following static curves induces a map

$$i_{t,\bar{t}} : (Z_t,(x,t)) \supset (B(x,t,\beta(t)^{-1}),(x,t)) \to (Z_{\bar{t}}, \gamma(\bar{t})) \tag{90.16}$$

satisfying

$$\| (i_{t,\bar{t}})^* g_{Z_{\bar{t}}} - g_{Z_t} \|_{C^{\beta(t)^{-1}}} < \beta(t). \tag{90.17}$$

Proof. This follows immediately from Lemma 88.1. □

Proof of Proposition 90.1. If for some $w > 0$ we have $\Lambda_w = \emptyset$ then $M^+(w,t) = \emptyset$ for large t. Thus if $\Lambda_w = \emptyset$ for all $w > 0$, we can take the empty collection of pointed hyperbolic manifolds and then 1 and 2 will be satisfied vacuously, and $\alpha(t)$ may be chosen so that 3 holds. So we assume that $\Lambda_w \neq \emptyset$ for some $w > 0$.

Since every complete finite-volume hyperbolic 3-manifold has a point with injectivity radius $\geq \mu_0$, there is a $w_0 > 0$ such that the collections $\{\Lambda_w\}_{w \leq w_0}$ contain the same sets of underlying hyperbolic manifolds (although the basepoints have more freedom when w is small). We let H_1 be a hyperbolic manifold from this collection with the fewest cusps and we choose a basepoint $x_1 \in H_1$ so that $(H_1,x_1) \in \Lambda_{w_0}$. Put $w_1 = \frac{w_0}{2}$. Note that x_1 lies in the w_1-thick part of H_1. In what follows we will use the fact that if f is a ϵ-approximation from H_1, for sufficiently small ϵ, then $f(x_1)$ will lie in the $.9w_0$-thick part of the image.

The idea of the first step of the proof is to define a family $\{f_0(t)\}$ of δ-approximations $(H_1,x_1) \to Z_t$, for all t sufficiently large, by taking a δ-approximation $(H_1,x_1) \to Z_t$, pushing it along static curves, and arguing using Lemma 90.11 that one can make small

adjustments from time to time to keep it a δ-approximation. The family $\{f_0(t)\}$ will not vary continuously with time, but it will have controlled "jumps".

More precisely, pick $T_0 < \infty$ and let $\xi_1, \ldots, \xi_4 > 0$ be parameters to be specified later. We assume that T_0 is large enough so that $2\beta(T_0) < \xi_1$, where β is from Lemma 90.14. By the definition of Λ_{w_1}, we may pick T_0 so that there is a point $(\overline{x}_0, T_0) \in M^+(w_1, T_0) \subset \mathcal{M}^+_{T_0}$ and a ξ_1-approximation $f_0(T_0) : (H_1, x_1) \to (Z_{T_0}, \overline{x}_0)$.

To do the induction step, for a given $j \geq 0$ suppose that at time $2^j T_0$ there is a point $(\overline{x}_j, 2^j T_0) \in M^+(w_1, 2^j T_0) \subset \mathcal{M}^+_{2^j T_0}$ and a ξ_1-approximation $f_0(2^j T_0) : (H_1, x_1) \to (Z_{2^j T_0}, \overline{x}_j)$. As mentioned above, if ξ_1 is small then in fact $\overline{x}_j \in M^+(.9w_0, 2^j T_0)$.

By part 2 of Lemma 90.14, provided that T_0 is sufficiently large we may define, for all $t \in [2^j T_0, 2^{j+1} T_0]$, a $2\xi_1$-approximation $f_0(t) : (H_1, x_1) \to Z_t$ by moving $f_0(2^j T_0)$ along static curves. Provided that ξ_1 is sufficiently small we will have $f_0(2^{j+1}T_0)(x_1) \in M^+(w_1, 2^{j+1}T_0)$ and then part 1 of Lemma 90.14 says there is some $(H', x') \in \Lambda_{w_1}$ with a $\beta(2^{j+1}T_0)$-approximation $\phi : (H', x') \to (Z_{2^{j+1}T_0}, f_0(2^{j+1}T_0)(x_1))$. Provided that $\beta(2^{j+1}T_0)$ and ξ_1 are sufficiently small, the partially defined map $\phi^{-1} \circ f_0(2^{j+1}T_0)$ will define a ξ_2-approximation from (H_1, x_1) to (H', x'). Hence provided that ξ_2 is sufficiently small, by Lemma 90.11 the map will be ξ_3-close to an isometry $\psi : (H_1, x_1) \to H'$. (In applying Lemma 90.11 we use the fact that H_1 is also a manifold with the fewest number of cusps in Λ_{w_1}.) Put $\phi_1 = \phi \circ \psi$. Provided that ξ_3 is sufficiently small, $f_0(2^{j+1}T_0)$ and ϕ_1 will be ξ_4-close as maps from (H_1, x_1) to $Z_{2^{j+1}T_0}$. Since ϕ_1 is a $\beta(2^{j+1}T_0)$-approximation precomposed with an isometry which shifts basepoints a distance at most ξ_3, it will be a $2\beta(2^{j+1}T_0)$-approximation provided that $\xi_3 < 1$ and $\beta(2^{j+1}T_0) < \frac{1}{2}$. We now redefine $f_0(2^{j+1}T_0)$ to be ϕ_1 and let $\overline{x}_{j+1}$ be the image of x_1 under ϕ_1. This completes the induction step.

In this way we define a family of partially defined maps $\{f_0(t) : (H_1, x_1) \to Z_t\}_{t \in [T_0, \infty)}$. From the construction, $f_0(2^j T_0)$ is a $2\beta(2^j T_0)$-approximation for all $j \geq 0$. Lemma 90.14 then implies that there is a function $\alpha_1 : [T_0, \infty) \to (0, \infty)$ decreasing to zero at infinity such that for all $t \in [T_0, \infty)$, $f_0(t)$ is an $\alpha_1(t)$-approximation, and for every $\bar{t} \in [t, 2t]$ we may slide $f_0(t)$ along static curves to define an $\alpha_1(t)$-approximation $h(\bar{t}) : (H_1, x_1) \to Z_t$ which is $\alpha_1(t)$-close to $f_0(\bar{t})$.

One may now employ a standard smoothing argument to convert the family $\{f_0(t)\}_{t \in [T_0, \infty)}$ into a family $\{f_1(t)\}_{t \in [T_0, \infty)}$ which satisfies the first two conditions of the proposition. If condition 3 fails to hold then we redefine the Λ_w's by considering limits of only those $\{(\mathcal{M}^+_{t_i}, (x_i, t_i), t_i^{-1} g(t_i))\}_{i=1}^\infty$ with $t_i \to \infty$ and $x_i \in M^+(w, t_i) \subset \mathcal{M}^+_{t_i}$ not in the image of $f_1(t_i)$. Repeating the construction we obtain a pointed hyperbolic manifold (H_2, x_2) and a family $\{f_2(t)\}$ defined for large t satisfying conditions 1 and 2, where $\operatorname{im}(f_2(t))$ is disjoint from $\operatorname{im}(f_1(t))$ for large t. Iteration of this procedure must stop after k steps for some finite number k, in view of the fact that $\overline{V} < \infty$ and the fact that there is a positive lower bound on the volumes of complete hyperbolic 3-manifolds. We get the desired family $\{f(t)\}$ by taking the union of the maps $f_1(t), \ldots, f_k(t)$.

91. Incompressibility of cuspidal tori

By Proposition 90.1, we know that for large times the thick part of the manifold can be parametrized by a collection of (truncated) finite volume hyperbolic manifolds. In this section we show that each cuspidal torus maps to an embedded incompressible torus in $\mathcal{M}_t$. (An alternative argument is given in Section 93.) The strategy, due to Hamilton, is to argue by contradiction. If such a torus were compressible then there would be an embedded compressing disk of least area at each time. By estimating the rate of change of the area of such disks one concludes that the area must go to zero in finite time, which is absurd.

Let T_0, α, $\{(H_1, x_1), \ldots, (H_k, x_k)\}$, B_t, and $f(t)$ be as in Proposition 90.1. We will consider a fixed H_i, with $1 \le i \le k$, which is noncompact. Choose a number $a > 0$ much smaller than the Margulis constant and let $\{V_1, \ldots, V_l\} \subset H_i$ be the cusp regions bounded by tori of diameter a. Each V_j is an embedded 3-dimensional submanifold (with boundary) of H_i and is isometric to the quotient of a horoball in hyperbolic 3-space $\mathbb{H}^3$ by the action of a copy of $\mathbb{Z}^2$ sitting in the stabilizer of the horoball. The boundary ∂V_j is a totally umbilic torus whose principal curvatures are equal to $\frac{1}{2}$ everywhere. We let $Y \subset H_i$ be the closure of the complement of $\bigcup_{j=1}^{l} V_j$ in H_i.

Let $T_a < \infty$ be large enough that B_{T_a} (defined as in Proposition 90.1) contains Y.

In order to focus on a given cusp, we now fix an integer $1 \le j \le l$ and put

$$Z = \partial V_j, \quad \hat{Z}_t = f(t)(Z) \quad \hat{Y}_t = f(t)(Y), \quad \hat{W}_t = \mathcal{M}_t^+ - \operatorname{int}(\hat{Y}_t) \tag{91.1}$$

for every $t \ge T_a$. The objective of this section is:

Proposition 91.2. *The homomorphism*

$$\pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\mathcal{M}_t^+, f(t)(\star)) \tag{91.3}$$

is a monomorphism for all $t \ge T_a$.

Proof. The proof will occupy the remainder of this section. The first step is:

Lemma 91.4. *The kernels of the homomorphisms*

$$\pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\mathcal{M}_t^+, f(t)(\star)), \quad \pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\hat{W}_t, f(t)(\star)) \tag{91.5}$$

are independent of t, for all $t \ge T_a$.

Proof. We prove the assertion for the first homomorphism. The argument for the second one is similar.

The kernel obviously remains constant on any time interval which is free of singular times. Suppose that $t_0 \ge T_a$ is a singular time. Then the intersection $\mathcal{M}_{t_0}^+ \cap \mathcal{M}_{t_0}^-$ includes into $\mathcal{M}_{t_0}^+$ and, by using static curves, into $\mathcal{M}_t$ for $t \neq t_0$ close to t_0. By Van Kampen's theorem, these inclusions induce monomorphisms of the fundamental groups. Therefore for t close to t_0, the kernel of (91.5) is the same as the kernel of

$$\pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\mathcal{M}_{t_0}^+ \cap \mathcal{M}_{t_0}^-), \tag{91.6}$$

which is independent of t for times t close to t_0. ∎

We now assume that the kernel of

$$\pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\mathcal{M}_t^+, f(t)(\star)) \tag{91.7}$$

is nontrivial for some, and hence every, $t \geq T_a$. By Van Kampen's theorem and the fact that the cuspidal torus $Z \subset Y$ is incompressible in Y, it follows that the kernel K of

$$\pi_1(f(t)) : \pi_1(Z, \star) \to \pi_1(\hat{W}_t, f(t)(\star)) \tag{91.8}$$

in nontrivial for all $t \geq T_a$. By Poincaré duality, $\mathrm{Im}\left(\mathrm{H}^1(\hat{W}_t;\mathbb{R}) \to \mathrm{H}^1(\partial\hat{W}_t;\mathbb{R})\right)$ is a Lagrangian subspace of $\mathrm{H}^1(\partial\hat{W}_t;\mathbb{R})$. In particular, $\mathrm{Im}\left(\mathrm{H}^1(\hat{W}_t;\mathbb{R}) \to \mathrm{H}^1(Z;\mathbb{R})\right)$ has rank one. Dually, $\mathrm{Ker}\left(\mathrm{H}_1(Z;\mathbb{R}) \to \mathrm{H}_1(\widehat{W_t};\mathbb{R})\right)$ has rank one and so K, a subgroup of a rank-two free abelian group, has rank one. We note that for all large t, $\hat{Z}_t$ is a convex boundary component of $\hat{W}_t$. The main theorem of [43] implies that for every such t, there is is a least-area compressing disk

$$(N_t^2, \partial N_t^2) \subset (\hat{W}_t, \hat{Z}_t). \tag{91.9}$$

We recall that a compressing disk is an embedded disk whose boundary curve is essential in $\hat{Z}_t$. We note that by definition, $\hat{W}_t$ is a compact manifold even when t is a singular time. The embedded curve $f(t)^{-1}(\partial N_t) \subset Z$ represents a primitive element of $\pi_1(Z)$ which, since K has rank one, must therefore generate K. It follows that modulo taking inverses, the homotopy class of $f(t)^{-1}(\partial N_t) \subset Z$ is independent of t.

We define a function $A : [T_a, \infty) \to (0, \infty)$ by letting $A(t)$ be the infimum of the areas of such embedded compressing disks. We now show that the least-area compressing disks avoid the surgery regions.

Lemma 91.10. *Let $\delta(t)$ be the surgery parameter from Section 73. There is a $T = T(a) < \infty$ so that whenever $t \geq T$, no point in any area-minimizing compressing disk $N_t \subset \hat{W}_t$ is in the center of a $10\delta(t)$-neck.*

Proof. If the lemma were not true then there would be a sequence of times $t_k \to \infty$ and for each k an area-minimizing compressing disk $(N_{t_k}, \partial N_{t_k}) \subset (\hat{W}_{t_k}, \hat{Z}_{t_k})$, along with a point $x_k \in N_{t_k}$ that is in the center of a $10\delta(t_k)$-neck. Note that the scalar curvature near ∂N_{t_k} is comparable to $-\frac{3}{2t_k}$. We now rescale by $R(x_k, t_k)$, and consider the map of pointed manifolds $f_k : (N_{t_k}, \partial N_{t_k}, x_k) \hookrightarrow (\hat{W}_{t_k}, \hat{Z}_{t_k}, x_k)$ where the domain is equipped with the pullback Riemannian metric. By [57] and standard elliptic regularity, for all $\rho < \infty$ and every integer j, the j^{th} covariant derivative of the second fundamental form of f_k is uniformly bounded on the ball $B(x_k, \rho) \subset N_{t_k}$, for sufficiently large k. Therefore the pointed Riemannian manifolds $(N_{t_k}, \partial N_{t_k}, x_k)$ subconverge in the smooth topology to a pointed, complete, connected, smooth manifold (N_∞, x_∞). Using the same bounds on the derivatives of the second fundamental form, we may extract a limit mapping $\phi_\infty : N_\infty \to \mathbb{R} \times S^2$ which is a 2-sided isometric stable minimal immersion. By [58, Theorem 2], ϕ_∞ is a totally geodesic immersion whose normal vector field in M has vanishing Ricci curvature. It follows that ϕ_∞ is a cover of a fiber $\{\mathrm{pt}\} \times S^2$. This contradicts the fact that N_∞ is noncompact. □

We redefine T_a if necessary so that T_a is greater than the T of Lemma 91.10.

We can isotope the surface Z by moving it down the cusp V_j. In doing so we do not change the group K but we can make the diameter of Z as small as desired. The next lemma refers to this isotopy freedom.

Lemma 91.11. *Given $D > 0$, there is a number $a_0 > 0$ so that for any $a \in (0, a_0)$, if* $\mathrm{diam}(Z) = a$ *and t is sufficiently large then $\int_{\partial N_t} \kappa_{\partial N_t} ds \le \frac{D}{2}$ and* $\mathrm{length}(\partial N_t) \le \frac{D}{2}\sqrt{t}$, *where $\kappa_{\partial N_t}$ is the geodesic curvature of $\partial N_t \subset N_t$.*

Proof. This is proved in [34, Sections 11 and 12]. We just state the main idea. For the purposes of this proof, we give $\hat{W}_t$ the metric $t^{-1}g(t)$. First, ∂N_t is the intersection of N_t with $\hat{Z}_t$. Because N_t is minimal with respect to free boundary conditions (i.e. the only constraint is that ∂N_t is in the right homotopy class in $\hat{Z}_t$), it follows that N_t meets $\hat{Z}_t$ orthogonally. Then $\kappa_{\partial N_t} = \Pi(v, v)$, where Π is the second fundamental form of $\hat{Z}_t$ in $\hat{W}_t$ and v is the unit tangent vector of ∂N_t. Given $a > 0$, let Z be the horospherical torus in V_j of diameter a. By Proposition 90.1, for large t the map $f(t)$ is close to being an isometry of pairs $(Y, Z) \to (\hat{Y}_t, \hat{Z}_t)$. As Z has principal curvatures $\frac{1}{2}$, we may assume that $\Pi(v, v)$ is close to $\frac{1}{2}$. This reduces the problem to showing that with an appropriate choice of a_0, if $a \in (0, a_0)$ then for large values of t the length of ∂N_t is guaranteed to be small. The intuition is that since $\hat{W}_t$ is close to being the standard cusp V_j, a large piece of the minimal disk N_t should be like a minimal surface N_∞ in V_j that intersects Z in the given homotopy class. Such a minimal surface in V_j essentially consists of a geodesic curve in Z going all the way down the cusp. The length of the intersection of N_∞ with the horospherical torus of diameter a is proportionate to a. Hence if a_0 is small enough, one would expect that if $a < a_0$ and if t is large then the length of ∂N_t is small. In particular, the length of ∂N_t is uniformly bounded with respect to a. A detailed proof appears in [34, Section 12].

Rescaling from the metric $t^{-1}g(t)$ to the original metric $g(t)$, $\int_{\partial N_t} \kappa_{\partial N_t} ds$ is unchanged and $\mathrm{length}(\partial N_t)$ is multiplied by $\sqrt{t}$. □

Lemma 91.12. *For every $D > 0$ there is a number $a_0 > 0$ with the following property. Given $a \in (0, a_0)$, suppose that we take Z to be the torus cross-section in V_j of diameter a. Then there is a number $T'_a < \infty$ so that as long as $t_0 \ge T'_a$, there is a smooth function $\bar{A}$ defined on a neighborhood of t_0 such that $\bar{A}(t_0) = A(t_0)$, $\bar{A} \ge A$ everywhere, and*

$$\bar{A}'(t_0) < \frac{3}{4}\left(\frac{1}{t_0 + \frac{1}{4}}\right) A(t_0) - 2\pi + D. \tag{91.13}$$

Proof. Take a_0 as in Lemma 91.11.

For $t_0 > T_a$, we begin with the minimizing compressing disk $N_{t_0} \subset \mathcal{M}^+_{t_0}$. If t_0 is a surgery time and N_{t_0} intersected the surgery region $\mathcal{M}^+_{t_0} - (\mathcal{M}^+_{t_0} \cap \mathcal{M}^-_{t_0})$ then N_{t_0} would have to pass through a $10\delta(t_0)$-neck, which is impossible by Lemma 91.10. Thus N_{t_0} avoids any parts added by surgery.

For t close to t_0 we define an embedded compressing disk $S_t \subset \mathcal{M}^+_t$ as follows. We take N_{t_0} and extend it slightly to a smooth surface $N'_{t_0} \subset \mathcal{M}^+_{t_0}$ which contains N_{t_0} in its interior. The surface N'_{t_0} will be unscathed on some open time interval containing t_0. If we let $S'_t \subset \mathcal{M}^+_t$ be the surface obtained by moving N'_{t_0} along static curves then for some $b > 0$, the surface

S'_t will intersect $\partial \hat{W}_t$ transversely for all $t \in (t_0 - b, t_0 + b)$. Putting

$$S_t = S'_t \cap \hat{W}_t \tag{91.14}$$

defines a compressing disk for $\hat{Z}_t \subset \hat{W}_t$.

Define $\bar{A} : (t_0 - b, t_0 + b) \to \mathbb{R}$ by

$$\bar{A}(t) = \mathrm{area}(S_t). \tag{91.15}$$

Clearly $\bar{A}(t_0) = A(t_0)$ and $\bar{A} \geq A$.

For the rest of the calculation, we will view S_t as a surface sitting in a fixed manifold M (a fattening of $\hat{W}_t$) with a varying metric $g(\cdot)$, and put $S = S_{t_0} = N_{t_0}$.

By the first variation formula for area,

$$\bar{A}'(t_0) = \int_{\partial S} \langle X, \nu_{\partial S} \rangle \, ds + \int_S \frac{d}{dt}\Big|_{t=t_0} \mathrm{dvol}_S, \tag{91.16}$$

where X denotes the variation vector field for $\hat{Z}_t$, viewed as a surface moving in M, and $\nu_{\partial S}$ is the outward normal vector along ∂S. By Proposition 90.1, there is an estimate $|X| \leq \alpha(t_0) t_0^{-\frac{1}{2}}$, where $\alpha(t_0) \to 0$ as $t_0 \to \infty$. Therefore

$$\left| \int_{\partial S_{t_0}} \langle X, \nu_{\partial S_{t_0}} \rangle \, ds \right| \leq \alpha(t_0) \, t_0^{-\frac{1}{2}} \; \mathrm{length}(\partial N_{t_0}). \tag{91.17}$$

By Lemma 91.11, the right-hand side of (91.17) is bounded above by $\frac{D}{2}$ if t_0 is large.

We turn to the second term in (91.16). Pick $p \in S$ and let e_1, e_2, e_3 be an orthonormal basis for T_pM with e_1 and e_2 tangent to S. Then

$$\frac{1}{\mathrm{dvol}_S} \frac{d}{dt}\Big|_{t=t_0} \mathrm{dvol}_S = \frac{1}{2} \sum_{i=1}^{2} \frac{dg}{dt}\Big|_{t=t_0} (e_i, e_i) = - \mathrm{Ric}(e_1, e_1) - \mathrm{Ric}(e_2, e_2). \tag{91.18}$$

Now

$$\begin{aligned} - \mathrm{Ric}(e_1, e_1) - \mathrm{Ric}(e_2, e_2)) &= -R + \mathrm{Ric}(e_3, e_3) = -R + K(e_3, e_1) + K(e_3, e_2) \\ &= -\frac{R}{2} - K(e_1, e_2) = -\frac{R}{2} - K_S + \mathrm{GK}_S, \end{aligned} \tag{91.19}$$

where K_S denotes the Gauss curvature of S and GK_S denotes the product of the principal curvatures. Applying the Gauss-Bonnet formula

$$\int_{\partial S} \kappa_{\partial S} \, ds = 2\pi - \int_S K_S \, \mathrm{vol}_S, \tag{91.20}$$

the fact that $\mathrm{GK}_S \leq 0$ (since S is time-t_0 minimal) and the inequality

$$R_{\min}(t) \geq -\frac{3}{2} \left(\frac{1}{t + \frac{1}{4}} \right) \tag{91.21}$$

from Lemma 79.11, we obtain

$$\int_S \frac{d}{dt}\Big|_{t=t_0} \mathrm{dvol}_S \leq \int_S \frac{3}{4} \left(\frac{1}{t_0 + \frac{1}{4}} \right) \mathrm{dvol}_S + \int_{\partial S} \kappa_{\partial S} ds - 2\pi. \tag{91.22}$$

By Lemma 91.11, if $a \in (0, a_0)$, $\mathrm{diam}(Z) = a$ and t_0 is sufficiently large then $\int_{\partial S} \kappa_{\partial S} ds \leq \frac{D}{2}$. Using (91.16), (91.17) and (91.22), if t_0 is large then

$$\bar{A}'(t_0) < \frac{3}{4} \left(\frac{1}{t_0 + \frac{1}{4}} \right) A(t_0) - 2\pi + D. \tag{91.23}$$

This proves the lemma. □

Proof of Proposition 91.2. Pick $D < 2\pi$. Let $a < a_0$ and T'_a be as in Lemma 91.12.

By Lemma 91.12, A is bounded on compact subsets of $[T'_a, \infty)$. By Lemma 91.10, for any $t \in [T'_a, \infty)$ we can find a compact set $K_t \subset \mathcal{M}^+_t$ so that for all $t' \in [T'_a, \infty)$ sufficiently close to t, the compressing disk $N_{t'}$ lies in K_t. If $(K_t, g(t))$ and $(K_t, g(t'))$ are e^σ-biLipschitz equivalent then $A(t) \leq e^{2\sigma} \,\mathrm{area}\,(N_{t'}) = e^{2\sigma}\, A(t')$ and $A(t') \leq e^{2\sigma}\,\mathrm{area}\,(N_t) = e^{2\sigma}\, A(t)$. It follows that A is continuous on $[T'_a, \infty)$.

For $t \geq T'_a$, put

$$F(t) = \left(t + \frac{1}{4} \right)^{-\frac{3}{4}} A(t) + 4(2\pi - D) \left(t + \frac{1}{4} \right)^{\frac{1}{4}}. \tag{91.24}$$

We claim that $F(t) \leq F(T'_a)$ for all $t \geq T'_a$. Suppose not. Put $t_0 = \inf\{t \geq T'_a \,:\, F(t) > F(T'_a)\}$. By continuity, $F(t_0) = F(T'_a)$. Consider the function $\bar{A}$ of Lemma 91.12. Put

$$\bar{F}(t) = \left(t + \frac{1}{4} \right)^{-\frac{3}{4}} \bar{A}(t) + 4(2\pi - D) \left(t + \frac{1}{4} \right)^{\frac{1}{4}}. \tag{91.25}$$

Then $\bar{F}(t_0) = F(t_0)$ and in a small interval around t_0, we have $\bar{F} \geq F$. However, (91.13) implies that $\bar{F}'(t_0) < 0$. There is some $\sigma > 0$ so that for $t \in (t_0, t_0 + \sigma)$, we have

$$F(t) \leq \bar{F}(t) \leq \bar{F}(t_0) + \frac{1}{2} \bar{F}'(t_0)\,(t - t_0) < \bar{F}(t_0) = F(t_0) = F(T'_a), \tag{91.26}$$

which contradicts the definition of t_0.

Thus if $t \geq T'_a$ then $F(t) \leq F(T'_a)$. This implies that $A(t)$ is negative for large t, which contradicts the fact that an area is nonnegative.

We have shown that the homomorphism (91.3) is injective if t is sufficiently large. In view of Lemma 91.4, the same statement holds for all $t \geq T_a$. □

92. II.7.4. The thin part is a graph manifold

This section is concerned with showing that the thin part $M^-(w, t)$ is a graph manifold. We refer to Appendix I for the definition of a graph manifold. We remind the reader that this completes the proof the geometrization conjecture.

The next two theorems are purely Riemannian. They say that if a 3-manifold is locally volume-collapsed, with sectional curvature bounded below, then it is a graph manifold. They differ slightly in their hypotheses.

Theorem 92.1. *(cf. Theorem II.7.4) Suppose that (M^α, g^α) is a sequence of compact oriented Riemannian 3-manifolds, closed or with convex boundary, and $w^\alpha \to 0$. Assume that*

(1) for each point $x \in M^\alpha$ there exists a radius $\rho = \rho^\alpha(x)$ not exceeding the diameter of M^α such that the ball $B(x, \rho)$ in the metric g^α has volume at most $w^\alpha \rho^3$ and sectional curvatures at least $-\rho^{-2}$.
(2) each component of the boundary of M^α has diameter at most w^α, and has a (topologically trivial) collar of length one, where the sectional curvatures are between $-\frac{1}{4} - \epsilon$ and $-\frac{1}{4} - \epsilon$.

Then for large α, M^α is diffeomorphic to a graph manifold.

Remark 92.2. A proof of Theorem 92.1 appears in [64, Section 8]. The proof in [64] is for closed manifolds, but in view of condition (2) the method of proof clearly goes through to manifolds-with-boundary as considered in Theorem 92.1. The statement of the theorem in [52, Theorem II.7.4] also has a condition $\rho < 1$, which seems to be unnecessary.

Theorem 92.3. *(cf. Theorem II.7.4) Suppose that (M^α, g^α) is a sequence of compact oriented Riemannian 3-manifolds, closed or with convex boundary, and $w^\alpha \to 0$. Assume that*
(1) for each point $x \in M^\alpha$ there exists a radius $\rho = \rho^\alpha(x)$ such that the ball $B(x, \rho)$ in the metric g^α has volume at most $w^\alpha \rho^3$ and sectional curvatures at least $-\rho^{-2}$.
(2) each component of the boundary of M^α has diameter at most w^α, and has a (topologically trivial) collar of length one, where the sectional curvatures are between $-\frac{1}{4} - \epsilon$ and $-\frac{1}{4} - \epsilon$.
(3) for every $w' > 0$ and $m \in \{0, 1, \ldots, [\epsilon^{-1}]\}$, there exist $\overline{r} = \overline{r}(w') > 0$ and $K_m = K_m(w') < \infty$ such that for sufficiently large α, if $r \in (0, \overline{r}]$ and a ball $B(x, r)$ in the metric g^α has volume at least $w' r^3$ and sectional curvatures at least $-r^{-2}$ then $|\nabla^m \operatorname{Rm}|(x) \leq K_m r^{-m-2}$.

Then for large α, M^α is diffeomorphic to a graph manifold.

Remark 92.4. The statement of this theorem in [52, Theorem II.7.4] has the stronger assumption that (3) holds for all $m \geq 0$. In the application to the locally collapsing part of the Ricci flow, it is not clear that this stronger condition holds. However, one does get a bound on a large number of derivatives, which is good enough.

Remark 92.5. As pointed out in [52, Section 7.4], adding condition (3) simplifies the proof and allows one to avoid both Alexandrov spaces and Perelman's stability theorem. (A proof of Perelman's stability theorem appears in [38]). Proofs of Theorem 92.3 are in [9], [40] and [46].

Remark 92.6. Comparing Theorems 92.1 and 92.3, Theorem 92.1 has the extra assumption that $\rho^\alpha(x)$ does not exceed the diameter of M^α. Without this extra assumption, the Alexandrov space arguments could give that for large α, M^α is homeomorphic to a nonnegatively curved Alexandrov space [64, Theorem 1.1(2)]. This does not immediately imply that M^α is a graph manifold.

Remark 92.7. We give some simple examples where the collapsing theorems apply. Let (Σ, g_{hyp}) be a closed surface with the hyperbolic metric. Let $S^1(\mu)$ be a circle of length μ and consider the Ricci flow on $S^1 \times \Sigma$ with the initial metric $S^1(\mu) \times (\Sigma, c_0 \, g_{hyp})$. The Ricci flow solution at time t is $S^1(\mu) \times (\Sigma, (c_0 + 2t) g_{hyp})$. In the rest of this example we consider the rescaled metric $t^{-1} g(t)$. Its diameter goes like $O(t^0)$ and its sectional curvatures go like $O(t^0)$. If we take ρ to be a small constant then for large t, $B(x, t, \rho)$ is approximately a circle bundle over a ball in a hyperbolic surface of constant sectional curvature $-\frac{1}{2}$, with

circle lengths that go like $t^{-1/2}$. The sectional curvature on $B(x,t,\rho)$ is bounded below by $-\rho^{-2}$, and $\rho^{-3}\ \mathrm{vol}(B(x,t,\rho)) \sim t^{-1/2}$. Theorems 92.1 and 92.3 both apply.

Next, consider a compact 3-dimensional nilmanifold that evolves under the Ricci flow. Let $\theta_1, \theta_2, \theta_3$ be affine-parallel 1-forms on M which lift to Maurer-Cartan forms on the Heisenberg group, with $d\theta_1 = d\theta_2 = 0$ and $d\theta_3 = \theta_1 \wedge \theta_2$. Consider the metric

$$g(t) = \alpha^2(t)\,\theta_1^2 + \beta^2(t)\,\theta_2^2 + \gamma^2(t)\,\theta_3^2. \tag{92.8}$$

Its sectional curvatures are $R_{1212} = -\frac{3}{4}\frac{\gamma^2}{\alpha^2\beta^2}$ and $R_{1313} = R_{2323} = \frac{1}{4}\frac{\gamma^2}{\alpha^2\beta^2}$. The Ricci tensor is

$$\mathrm{Ric} = \frac{1}{2}\frac{\gamma^2}{\alpha^2\beta^2}\left(-\alpha^2\,\theta_1^2 - \beta^2\,\theta_2^2 + \gamma^2\,\theta_3^2\right). \tag{92.9}$$

The general solution to the Ricci flow equation is of the form

$$\begin{aligned} \alpha^2(t) &= A_0(t+t_0)^{1/3}, \\ \beta^2(t) &= B_0(t+t_0)^{1/3}, \\ \gamma^2(t) &= \frac{A_0B_0}{3}(t+t_0)^{-1/3}. \end{aligned} \tag{92.10}$$

In the rest of this example we consider the rescaled metric $t^{-1}g(t)$. Its diameter goes like $t^{-1/3}$, its volume goes like $t^{-4/3}$ and its sectional curvatures go like t^0. If we take $\rho(x) = \mathrm{diam}$ then $\rho^{-3}\,\mathrm{vol}(B(x,\rho)) \sim t^{-1/3}$, so both Theorem 92.1 and Theorem 92.3 apply. We could also take $\rho(x)$ to be a small constant $c > 0$, in which case Theorem 92.3 applies.

In general, among the eight maximal homogeneous geometries, the rescaled solution for a compact 3-manifold with geometry $H^2 \times \mathbb{R}$ or $\widetilde{\mathrm{SL}_2(\mathbb{R})}$ will collapse to a hyperbolic surface of constant sectional curvature $-\frac{1}{2}$. The rescaled solution for a Sol geometry will collapse to a circle. The rescaled solution for an $\mathbb{R}^3$ or Nil geometry will collapse to a point.

We remark that although these homogeneous solutions are collapsing in the sense of Theorem 92.1, there is no contradiction with the no local collapsing result of Theorem 26.2, which only rules out local collapsing on a finite time interval.

Returning to our Ricci flow with surgery, recall the statement of Proposition 90.1. If the collection $\{H_1, \ldots, H_k\}$ of Proposition 90.1 is nonempty then for large t, let $\widehat{H}_i(t)$ be the result of removing from H_i the horoballs whose boundaries are at distance approximately $\frac{1}{2}\,\alpha(t)$ from the basepoint x_i. (If there are no such horoballs then $\widehat{H}_i(t) = H_i$.) Put

$$M_{thin}(t) = \mathcal{M}_t^+ - f(t)(\widehat{H}_1(t) \cup \ldots \widehat{H}_k(t)). \tag{92.11}$$

Proposition 92.12. *For large t, $M_{thin}(t)$ is a graph manifold.*

Proof. We give two closely related proofs, one using Theorem 92.3 and one using Theorem 92.1.

If the proposition is not true then there is a sequence $t^\alpha \to \infty$ so that for each α, $M_{thin}(t^\alpha)$ is not a graph manifold. Let M^α be the manifold obtained from $M_{thin}(t^\alpha)$ by throwing away connected components which are closed and admit metrics of nonnegative sectional curvature, and put $g^\alpha = (t^\alpha)^{-1}g(t^\alpha)$. Since any closed manifold of nonnegative

sectional curvature is a graph manifold by [35], for each α the manifold M^α is not a graph manifold.

We first show that the assumptions of Theorem 92.3 are verified.

Lemma 92.13. *Condition (3) in Theorem 92.3 holds for the M^α's.*

Proof. With w' being a parameter as in Condition (3) in Theorem 92.3, let $\overline{r} = \overline{r}(w')$ be the parameter of Corollary 81.3. It is enough to show that for large α, if $r \in (0, \overline{r}\sqrt{t^\alpha}]$, $x^\alpha \in \mathcal{M}^+_{t^\alpha}$, and $B(x^\alpha, t^\alpha, r)$ has volume at least $w' r^3$ and sectional curvatures bounded below by $-r^{-2}$, then $|\nabla^m \operatorname{Rm}|(x^\alpha, t^\alpha) \leq K_m\, r^{-m-2}$ for an appropriate choice of constants K_m.

To prove this by contradiction, we assume that after passing to a subsequence if necessary, there are $r^\alpha \in (0, \overline{r}\sqrt{t^\alpha}]$ and $x^\alpha \in \mathcal{M}^+_{t^\alpha}$ such that $B(x^\alpha, t^\alpha, r^\alpha)$ has volume at least $w'(r^\alpha)^3$ and sectional curvature at least $-(r^\alpha)^{-2}$, but $\lim_{\alpha\to\infty} (r^\alpha)^{m+2}\, |\nabla^m \operatorname{Rm}|(x^\alpha, t^\alpha) = \infty$ for some $m \leq [\epsilon^{-1}]$.

In the notation of Corollary 81.3, if $r^\alpha \geq \theta^{-1}(w') h_{\max}(t^\alpha)$ for infinitely many α then for these α, Corollary 81.3 gives a curvature bound on an unscathed parabolic neighborhood $P\left(x^\alpha, t^\alpha, r^\alpha/4, -\tau(r^\alpha)^2\right)$ and hence, by Appendix D, derivative bounds at (x^α, t^α). This is a contradiction. Therefore we may assume that $r^\alpha < \theta^{-1}(w') h_{\max}(t^\alpha) \leq r(t^\alpha)$ for all α, where we used Remark 86.4 for the last inequality.

Suppose first that $R(x^\alpha, t^\alpha) \leq (r(t^\alpha))^{-2}$. By Lemma 70.1, there is an estimate $R \leq 16\,(r(t^\alpha))^{-2}$ on the parabolic neighborhood $P\left(x^\alpha, t^\alpha, \frac{1}{4}\eta^{-1} r(t^\alpha), -\frac{1}{16}\eta^{-1}\left(r(t^\alpha)\right)^2\right)$. A surgery in this neighborhood could only occur where $R \geq h_{\max}(t^\alpha)^{-2}$. For large α, $h_{\max}(t^\alpha)^{-2} >> r(t^\alpha)^{-2}$ by Remark 86.4. Hence this neighborhood is unscathed. Appendix D now gives bounds of the form $|\nabla^m \operatorname{Rm}|(x^\alpha, t^\alpha) \leq \text{const.}\ (r(t^\alpha))^{-m-2} \leq \text{const.}\ (r^\alpha)^{-m-2}$, which is a contradiction..

Suppose now that $R(x^\alpha, t^\alpha) > (r(t^\alpha))^{-2}$. Then (x^α, t^α) is in the center of a canonical neighborhood and there are universal estimates $|\nabla^m \operatorname{Rm}|(x^\alpha, t^\alpha) \leq \text{const.}(m)\, R(x^\alpha, t^\alpha)^{\frac{m+2}{2}}$ for all $m \leq [\epsilon^{-1}]$. Hence in this case, it suffices to show that $R(x^\alpha, t^\alpha)$ is bounded above by a constant times $(r^\alpha)^{-2}$, i.e. it suffices to get a contradiction just to the assumption that $\lim_{\alpha\to\infty} (r^\alpha)^2\, R(x^\alpha, t^\alpha) = \infty$.

So suppose that $\lim_{\alpha\to\infty} (r^\alpha)^2 R(x^\alpha, t^\alpha) = \infty$. We claim that $\lim_{\alpha\to\infty} (r^\alpha)^2 \inf R\Big|_{B(x^\alpha, t^\alpha, r^\alpha)} = \infty$. Suppose not. Then there is some $C \in (0, \infty)$ so that after passing to a subsequence, there are points $x^{\alpha\prime} \in B\left(x^\alpha, t^\alpha, r^\alpha\right)$ with $(r^\alpha)^2\, R(x^{\alpha\prime}, t^\alpha) < C$. Considering points along the time-t^α geodesic segment from $x^{\alpha\prime}$ to x^α, for large α we can find points $x^{\alpha\prime\prime} \in B\left(x^\alpha, t^\alpha, r^\alpha\right)$ with $(r^\alpha)^2\, R(x^{\alpha\prime\prime}, t^\alpha) = 2C$. Applying Lemma 70.2 at $(x^{\alpha\prime\prime}, t^\alpha)$, or more precisely a version that applies along geodesics as in Claim 2 of II.4.2, we obtain a contradiction to the assumption that $\lim_{\alpha\to\infty} (r^\alpha)^2\, R(x^\alpha, t^\alpha) = \infty$. In applying Lemma 70.2 we use that $r^\alpha < r(t^\alpha)$ and $\lim_{\alpha\to\infty} r(t^\alpha) = 0$, giving $\lim_{\alpha\to\infty} \left(r^\alpha\right)^{-2} t^\alpha = \infty$, in order to say that $\lim_{\alpha\to\infty} \frac{\Phi\left(2C(r^\alpha)^{-2}\right)}{2C(r^\alpha)^{-2}} = 0$.

Hence for large α, every point $x \in B\left(x^\alpha, t^\alpha, r^\alpha\right)$ is in the center of a canonical neighborhood of size comparable to $R(x, t^\alpha)^{-\frac{1}{2}}$, which is small compared to r^α. On the other hand,

from Lemma 83.1, there is a ball B' of radius $\theta_0(w')\, r^\alpha$ in $B\left(x^\alpha, t^\alpha, r^\alpha\right)$ so that every subball of B' has almost-Euclidean volume. This is a contradiction.

This proves the lemma. □

We continue with the proof of Proposition 92.12. By construction there is a sequence $w^\alpha \to 0$ so that conditions (1) and (2) of Theorem 92.3 hold with $\rho^\alpha(x^\alpha) = (t^\alpha)^{-\frac{1}{2}}\, \rho(x^\alpha, t^\alpha)$. Hence for large α, M^α is diffeomorphic to a graph manifold. This is a contradiction to the choice of the M^α's and proves the theorem.

We now give a proof that instead uses Theorem 92.1. Let d^α denote the diameter of M^α. If we take $\rho^\alpha(x^\alpha) = (t^\alpha)^{-\frac{1}{2}}\, \rho(x^\alpha, t^\alpha)$ then we can apply Theorem 92.1 as long as that the diameter statement in condition (1) of Theorem 92.1 is satisfied. If it is not satisfied then there is some point $x^\alpha \in M^\alpha$ with $\rho^\alpha(x^\alpha) > d^\alpha$. The sectional curvatures of M^α are bounded below by $-\frac{1}{\rho^\alpha(x^\alpha)^2}$, and so are bounded below by $-\frac{1}{(d^\alpha)^2}$. If there is a subsequence with $\frac{\rho^\alpha(x^\alpha)}{d^\alpha} \le C < \infty$ then

$$\mathrm{vol}(M^\alpha) \;=\; \mathrm{vol}(B(x^\alpha, \rho^\alpha(x^\alpha))) \;\le\; w^\alpha\, \rho^\alpha(x^\alpha)^3 \;\le\; w^\alpha\, C^3\, (d^\alpha)^3 \tag{92.14}$$

and we can apply Theorem 92.1 with $\rho^\alpha = d^\alpha$, after redefining w^α. Thus we may assume that $\lim_{\alpha\to\infty} \frac{\rho^\alpha(x^\alpha)}{d^\alpha} = \infty$. If there is a subsequence with $\frac{\mathrm{vol}(M^\alpha)}{(d^\alpha)^3} \to 0$ then we can apply Theorem 92.1 with $\rho^\alpha = d^\alpha$. Thus we may assume that $\frac{\mathrm{vol}(M^\alpha)}{(d^\alpha)^3}$ is bounded away from zero. After rescaling the metric to make the diameter one, we are in a noncollapsing situation with the lower sectional curvature bound going to zero. By the argument in the proof of Lemma 92.13 (which used Corollary 81.3) there are uniform L^∞-bounds on $\mathrm{Rm}(M^\alpha)$ and its covariant derivatives. After passing to a subsequence, there is a limit (M_∞, g_∞) in the smooth topology which is diffeomorphic to M^α for large α, and carries a metric of nonnegative sectional curvature. As any boundary component of M_∞ would have to have a neighborhood of negative sectional curvature (see the definition of M_{thin} and condition (2) of Theorem 92.1), M_∞ is closed. However, by construction M^α has no connected components which are closed and admit metrics of nonnegative sectional curvature. This contradiction shows that the diameter statement in condition (1) of Theorem 92.1 is satisfied. The other conditions of Theorem 92.1 are satisfied as before. □

Thus for large t, $\mathcal{M}_t^+$ has a decomposition into a piece $f(t)(\widehat{H}_1(t) \cup \ldots \widehat{H}_k(t))$, whose interior admits a complete finite-volume hyperbolic metric, and the complement, which is a graph manifold.

In addition, by Section 91 the cuspidal tori are incompressible in $\mathcal{M}_t^+$. By Lemma 73.4, the initial (connected) manifold $\mathcal{M}_0$ is diffeomorphic to a connected sum of the connected components of $\mathcal{M}_t$, along with some possible additional connected sums with a finite number of $S^1 \times S^2$'s and quotients of the round S^3. This proves the geometrization conjecture of Appendix I.

93. II.8. Alternative proof of cusp incompressibility

The goal of this section is to prove Perelman's Proposition II.8.2, which gives a numerical characterization of the geometric type of a compact 3-manifold. It also contains an independent proof of the incompressibility of the cuspidal ends of the hyperbolic piece in the geometric decomposition.

We recall from Section 7 that $\lambda(g)$ is the first eigenvalue of $-4\triangle + R$, and can also be expressed as

$$\lambda(g) \;=\; \inf_{\Phi \in C^\infty(M)\,:\,\Phi \neq 0} \frac{\int_M (4|\nabla \Phi|^2 \;+\; R\,\Phi^2)\; dV}{\int_M \Phi^2 \, dV}. \tag{93.1}$$

From Lemma 7.11, if $g(\cdot)$ is a Ricci flow and $\lambda(t) \;=\; \lambda(g(t))$ then

$$\frac{d}{dt}\lambda(t) \;\geq\; \frac{2}{3}\,\lambda^2(t). \tag{93.2}$$

From Lemma 8.1, $\lambda(t)\,V(t)^{\frac{2}{3}}$ is nondecreasing when it is nonpositive.

For any metric g, there are inequalities

$$\min R \;\leq\; \lambda(g) \;\leq\; \frac{\int_M R\,dV}{\mathrm{vol}(M,g)}, \tag{93.3}$$

where the first inequality follows directly from (93.1) and the second inequality comes from using 1 as a test function in (93.1).

Perelman's proof of his Proposition II.8.2 uses the functional $\lambda(g)V(g)^{\frac{2}{3}}$. The functional $R_{min}V(g)^{\frac{2}{3}}$ plays a similar role. For example, from Corollary 87.7, $\widehat{R}(t) \;=\; R_{min}(t)V(t)^{\frac{2}{3}}$ is nondecreasing when it is nonpositive. We first give a proof of an analog of Proposition II.8.2 that uses $R_{min}V(g)^{\frac{2}{3}}$ instead of $\lambda(g)V(g)^{\frac{2}{3}}$. The technical simplification is that when $R_{min}V(g)^{\frac{2}{3}}$ is nonpositive, it is nondecreasing under a surgery, as surgeries are only done in regions of large positive scalar curvature, so R_{min} doesn't change, and a surgery reduces volume. (A possible extinction of a component clearly doesn't change $R_{min}V(g)^{\frac{2}{3}}$.) We show that a minimal-volume hyperbolic submanifold of M has incompressible tori, which gives a different approach to Section 91.

Perelman's alternative approach to Section 91 uses the functional $\lambda(g)V(g)^{\frac{2}{3}}$ instead of $R_{min}V(g)^{\frac{2}{3}}$. Our use of $R_{min}V(g)^{\frac{2}{3}}$ and the sigma-invariant $\sigma(M)$, instead of $\lambda(g)V(g)^{\frac{2}{3}}$ and $\overline{\lambda}$, is inspired by [4].

We then give the arguments using $\lambda(g)V(g)^{\frac{2}{3}}$, thereby proving Perelman's Proposition II.8.2. The main technical difficulty is to control how $\lambda(g)V(g)^{\frac{2}{3}}$ changes under a surgery.

93.1. **The approach using the σ-invariant.** We first give some well-known results about the sigma-invariant. We recall that the sigma-invariant of a closed connected manifold M of dimension $n \geq 3$ is given by

$$\sigma(M) \;=\; \sup_{\mathcal{C}} \inf_{g \in \mathcal{C}} \frac{\int_M R(g)\; \mathrm{dvol}(g)}{\mathrm{vol}(M,g)^{\frac{n-2}{n}}}, \tag{93.4}$$

where $\mathcal{C}$ runs over the conformal classes of Riemannian metrics on M. From the solution to the Yamabe problem, the infimum in (93.4) is realized by a metric of constant scalar curvature in the given conformal class. It follows that if $\sigma(M) > 0$ then M admits a metric with positive scalar curvature. Conversely, suppose that M admits a metric g_0 with positive scalar curvature. Let $\mathcal{C}$ be the conformal class containing g_0. Then

$$\inf_{g \in \mathcal{C}} \frac{\int_M R(g)\ \mathrm{dvol}(g)}{\mathrm{vol}(M,g)^{\frac{n-2}{n}}} = \inf_{u>0} \frac{\int_M \left(\frac{4(n-1)}{n-2} \, |\nabla u|^2 + R(g_0)\, u^2 \right) \mathrm{dvol}_M(g_0)}{\left(\int_M u^{\frac{2n}{n-2}} \ \mathrm{dvol}_M(g_0) \right)^{\frac{n-2}{n}}} \tag{93.5}$$

is positive, in view of the Sobolev embedding theorem, and so $\sigma(M) > 0$.

We claim that if $\sigma(M) \leq 0$ then

$$\sigma(M) = \sup_g R_{min}(g)\, V(g)^{\frac{2}{n}}. \tag{93.6}$$

To see this, as the infimum in (93.4) is realized by a metric of constant scalar curvature in the given conformal class, it follows that $\sigma(M) \leq \sup_g R_{min}(g)\, V(g)^{\frac{2}{n}}$. Now given a Riemannian metric g, the infimum in (93.4) within the corresponding conformal class $\mathcal{C}$ equals $\widetilde{R}\, V(\widetilde{g})^{\frac{2}{n}}$ for a metric $\widetilde{g} = u^{\frac{4}{n-2}} g$ with constant scalar curvature $\widetilde{R}$. Then

$$\widetilde{R}\, V(\widetilde{g})^{\frac{2}{n}} = \inf_{u>0} \frac{\int_M \left(\frac{4(n-1)}{n-2} \, |\nabla u|^2 + R\, u^2 \right) \mathrm{dvol}_M}{\left(\int_M u^{\frac{2n}{n-2}} \ \mathrm{dvol}_M \right)^{\frac{n-2}{n}}}. \tag{93.7}$$

As

$$\frac{\int_M \left(\frac{4(n-1)}{n-2} \, |\nabla u|^2 + R\, u^2 \right) \mathrm{dvol}_M}{\left(\int_M u^{\frac{2n}{n-2}} \ \mathrm{dvol}_M \right)^{\frac{n-2}{n}}} \geq R_{min}(g) \frac{\int_M u^2 \ \mathrm{dvol}_M}{\left(\int_M u^{\frac{2n}{n-2}} \ \mathrm{dvol}_M \right)^{\frac{n-2}{n}}} \tag{93.8}$$

and $R_{min}(g) \leq 0$, Holder's inequality implies that $\widetilde{R}\, V(\widetilde{g})^{\frac{2}{n}} \geq R_{min}(g)\, V(g)^{\frac{2}{n}}$. It follows that $\sigma(M) \geq \sup_g R_{min}(g)\, V(g)^{\frac{2}{n}}$.

The next proposition answers conjectures of Anderson [2].

Proposition 93.9. *Let M be a closed connected oriented 3-manifold.*
(a) If $\sigma(M) > 0$ then M is diffeomorphic to a connected sum of a finite number of $S^1 \times S^2$'s and metric quotients of the round S^3. Conversely, each such manifold has $\sigma(M) > 0$.
(b) M is a graph manifold if and only if $\sigma(M) \geq 0$.
(c) If $\sigma(M) < 0$ then $\left(- \frac{2}{3}\, \sigma(M) \right)^{\frac{3}{2}}$ is the minimum of the numbers V with the following property : M can be decomposed as a connected sum of a finite collection of $S^1 \times S^2$'s, metric quotients of the round S^3 and some other components, the union of which is denoted by M', and there exists a (possibly disconnected) complete finite-volume manifold N with constant sectional curvature $-\frac{1}{4}$ and volume V which can be embedded in M' so that the complement $M' - N$ (if nonempty) is a graph manifold.

Moreover, if $\mathrm{vol}(N) = \left(- \frac{2}{3}\, \sigma(M) \right)^{\frac{3}{2}}$ *then the cusps of N (if any) are incompressible in M'.*

Proof. If $\sigma(M) > 0$ then M has a metric g of positive scalar curvature. From Lemmas 81.1 and 81.2, M is a connected sum of $S^1 \times S^2$'s and metric quotients of the round S^3. Conversely, if M is a connected sum of $S^1 \times S^2$'s and metric quotients of the round S^3 then M admits a metric g of positive scalar curvature and so $\sigma(M) > 0$.

Now suppose that $\sigma(M) \leq 0$. If M is a graph manifold then M volume-collapses with bounded curvature, so (93.6) implies that $\sigma(M) = 0$.

Suppose that M is not a graph manifold. Suppose that we have a given decomposition of M as a connected sum of a finite collection of $S^1 \times S^2$'s, metric quotients of the round S^3 and some other components, the union of which is denoted by M', and there exists a (possibly disconnected) finite-volume complete manifold N with constant sectional curvature $-\frac{1}{4}$ which can be embedded in M' so that the complement (if nonempty) is a graph manifold. Let V_{hyp} denote the hyperbolic volume of N. We do not assume that the cusps of N are incompressible in M'. For any $\epsilon > 0$, we claim that there is a metric g_ϵ on M with $R \geq -6 \cdot \frac{1}{4} - \epsilon$ and volume $V(g_\epsilon) \leq V_{hyp} + \epsilon$. This comes from collapsing the graph manifold pieces, along with the fact that the connected sum operation can be performed while decreasing the scalar curvature arbitrarily little and increasing the volume arbitrarily little. Then $R_{min}(g_\epsilon)\, V(g_\epsilon)^{\frac{2}{3}} \geq -\frac{3}{2}\, V_{hyp}^{2/3} - \text{const.}\, \epsilon$. Thus $\sigma(M) \geq -\frac{3}{2}\, V_{hyp}^{2/3}$.

Let $\widehat{V}$ denote the minimum of V_{hyp} over all such decompositions of M. (As the set of volumes of complete finite-volume 3-manifolds with constant curvature $-\frac{1}{4}$ is well-ordered, there is a minimum.) Then $\sigma(M) \geq -\frac{3}{2}\, \widehat{V}^{2/3}$.

Next, take an arbitrary metric g_0 on M and consider the Ricci flow $g(t)$ with initial metric g_0. From Sections 90 and 92, there is a nonempty manifold N with a complete finite-volume metric of constant curvature $-\frac{1}{4}$ so that for large t, there is a decomposition $\mathcal{M}_t^+ = M_1(t) \cup M_2(t)$ of the time-t manifold, where $M_1(t)$ is a graph manifold and $(M_2(t), \frac{1}{t} g(t)\big|_{M_2(t)})$ is close to a large piece of N. In terms of condition (c) of Proposition 93.9, we will think of M' as being $\mathcal{M}_t^+$. Because of the presence of N, we know that $t\, R_{min}(t) \leq -\frac{3}{2} + \epsilon(t)$ and $V(t) \geq t^{2/3}\, V_{hyp}(N) - \epsilon(t)$ for a function $\epsilon(t)$ with $\lim_{t \to \infty} \epsilon(t) = 0$. The monotonicity of $R_{min}(t)\, V^{2/3}(t)$, even through surgeries, implies that

$$R_{min}(g_0)\, V^{2/3}(g_0) \leq -\frac{3}{2}\, V_{hyp}(N)^{2/3} \leq -\frac{3}{2}\, \widehat{V}^{2/3}. \tag{93.10}$$

Thus $\sigma(M) \leq -\frac{3}{2}\, \widehat{V}^{2/3}$.

This shows that $\sigma(M) = -\frac{3}{2}\, \widehat{V}^{2/3}$. Now take a decomposition of M as in condition (c) of Proposition 93.9, with $V_{hyp}(N) = \widehat{V}$. We claim that the cuspidal 2-tori of N are incompressible in M'. If not then there would be a metric g on M with $R(g) \geq -\frac{3}{2}$ and $\mathrm{vol}(g) < V_{hyp}(N)$ [3, Pf. of Theorem 2.9]. This would contradict the fact that $\sigma(M) = -\frac{3}{2}\, V_{hyp}(N)^{2/3}$. □

93.2. **The approach using the $\overline{\lambda}$-invariant.**

Proposition 93.11. *(cf. II.8.2) Let M be a closed connected oriented 3-manifold.*
(a) If M admits a metric g with $\lambda(g) > 0$ then it is diffeomorphic to a connected sum of a finite number of $S^1 \times S^2$'s and metric quotients of the round S^3. Conversely, each such

manifold admits a metric g with $\lambda(g) > 0$.
(b) Suppose that M does not admit any metric g with $\lambda(g) > 0$. Let $\overline{\lambda}$ denote the supremum of $\lambda(g)V(g)^{\frac{2}{3}}$ over all metrics g on M. Then M is a graph manifold if and only if $\overline{\lambda} = 0$.
(c) Suppose that M does not admit any metric g with $\lambda(g) > 0$, and $\overline{\lambda} < 0$. Then $\left(-\frac{2}{3}\overline{\lambda}\right)^{\frac{3}{2}}$ is the minimum of the numbers V with the following property : M can be decomposed as a connected sum of a finite collection of $S^1 \times S^2$'s, metric quotients of the round S^3 and some other components, the union of which is denoted by M', and there exists a (possibly disconnected) complete manifold N with constant sectional curvature $-\frac{1}{4}$ and volume V which can be embedded in M' so that the complement $M' - N$ (if nonempty) is a graph manifold.

Moreover, if $\mathrm{vol}(N) = \left(-\frac{2}{3}\overline{\lambda}\right)^{\frac{3}{2}}$ then the cusps of N (if any) are incompressible in M'.

Proof. We first give the argument for Proposition 93.11 under the pretense that all Ricci flows are smooth, except for possible extinction of components. (Of course this is not the case, but it will allow us to present the main idea of the proof.)

If $\lambda(g) > 0$ for some metric g then from (93.2), the Ricci flow starting from g will become extinct within time $\frac{3}{2\lambda(g)}$. Hence Lemma 81.2 applies. Conversely, if M is a connected sum of $S^1 \times S^2$'s and metric quotients of the round S^3 then M admits a metric g of positive scalar curvature. From (93.3), $\lambda(g) > 0$.

Now suppose that M does not admit any metric g with $\lambda(g) > 0$. If M is a graph manifold then M volume-collapses with bounded curvature, so (93.3) implies that $\overline{\lambda} = 0$.

Suppose that M is not a graph manifold. Suppose that we have a given decomposition of M as a connected sum of a finite collection of $S^1 \times S^2$'s, metric quotients of the round S^3 and some other components, the union of which is denoted by M', and there exists a (possibly disconnected) complete manifold N with constant sectional curvature $-\frac{1}{4}$ which can be embedded in M' so that the complement (if nonempty) is a graph manifold. Let V_{hyp} denote the hyperbolic volume of N. We do not assume that the cusps of N are incompressible in M'. For any $\epsilon > 0$, we claim that there is a metric g_ϵ on M with $R \geq -6 \cdot \frac{1}{4} - \epsilon$ and volume $V(g_\epsilon) \leq V_{hyp} + \epsilon$. This comes from collapsing the graph manifold pieces, along with the fact that the connected sum operation can be performed while decreasing the scalar curvature arbitrarily little and increasing the volume arbitrarily little. Then (93.3) implies that $\lambda(g_\epsilon)\, V(g_\epsilon)^{\frac{2}{3}} \geq -\frac{3}{2}\, V_{hyp}^{2/3} - \text{const.}\ \epsilon$. Thus $\overline{\lambda} \geq -\frac{3}{2}\, V_{hyp}^{2/3}$.

Let $\widehat{V}$ denote the minimum of V_{hyp} over all such decompositions of M. (As the set of volumes of complete finite-volume 3-manifolds with constant curvature $-\frac{1}{4}$ is well-ordered, there is a minimum.) Then $\overline{\lambda} \geq -\frac{3}{2}\, \widehat{V}^{2/3}$.

Next, take an arbitrary metric g_0 on M and consider the Ricci flow $g(t)$ with initial metric g_0. From Sections 90 and 92, there is a nonempty manifold N with a finite-volume complete metric of constant curvature $-\frac{1}{4}$ so that for large t, there is a decomposition of the time-t manifold $\mathcal{M}_t^+ = M_1(t) \cup M_2(t)$ where $M_1(t)$ is a graph manifold and $(M_2(t), \frac{1}{t}\, g(t)\big|_{M_2(t)})$ is

close to a large piece of N. As N has finite volume, a constant function on N is square-integrable and so $\inf \operatorname{spec}(-\triangle_N) = 0$. Equivalently,

$$\inf_{f \in C_c^\infty(N), f \neq 0} \frac{\int_N |\nabla f|^2 \ \mathrm{dvol}_N}{\int_N f^2 \ \mathrm{dvol}_N} = 0. \tag{93.12}$$

Taking an appropriate test function Φ on M with compact support in $M_2(t)$ gives $t\,\lambda(t) \leq -\frac{3}{2}+\epsilon_1(t)$, with $\lim_{t\to\infty} \epsilon_1(t) = 0$. In terms of condition (c) of Proposition 93.11, we will think of M' as being $\mathcal{M}_t^+$. From the presence of N, we know that $V(t) \geq t^{2/3} V_{hyp}(N) - \epsilon_2(t)$, with $\lim_{t\to\infty} \epsilon_2(t) = 0$. As we are assuming that the Ricci flow is nonsingular, the monotonicity of $\lambda(t)\, V^{2/3}(t)$ implies that

$$\lambda(g_0)\, V^{2/3}(g_0) \leq -\frac{3}{2} V_{hyp}(N)^{2/3} \leq -\frac{3}{2} \widehat{V}^{2/3}. \tag{93.13}$$

Thus $\overline{\lambda} \leq -\frac{3}{2} \widehat{V}^{2/3}$.

This shows that $\overline{\lambda} = -\frac{3}{2} \widehat{V}^{2/3}$. Now take a decomposition of M as in condition (c) of Proposition 93.11, with $V_{hyp}(N) = \widehat{V}$. We claim that the cuspidal 2-tori of N are incompressible in M'. If not then there would be a metric g on M with $R(g) \geq -\frac{3}{2}$ and $\mathrm{vol}(g) < V_{hyp}(N)$ [3, Pf. of Theorem 2.9]. Using (93.3), one would obtain a contradiction to the fact that $\overline{\lambda} = -\frac{3}{2} V_{hyp}(N)^{2/3}$.

To handle the behaviour of $\lambda(t)\, V^{\frac{2}{3}}(t)$ under Ricci flows with surgery, we first state a couple of general facts about Schrödinger operators.

Lemma 93.14. *Given a closed Riemannian manifold M, let X be a codimension-0 submanifold-with-boundary of M. Given $R \in C^\infty(M)$, let λ_M be the lowest eigenvalue of $-4\triangle + R$ on M, with corresponding eigenfunction ψ. Let λ_X be the lowest eigenvalue of the corresponding operator on X, with Dirichlet boundary conditions, and similarly for $\lambda_{M-\mathrm{int}(X)}$. Then for all $\eta \in C_c^\infty(\mathrm{int}(X))$, we have*

$$\lambda_M \leq \min(\lambda_X, \lambda_{M-\mathrm{int}(X)}) \tag{93.15}$$

and

$$\lambda_X \leq \lambda_M + 4 \frac{\int_M |\nabla \eta|^2\, \psi^2\, dV}{\int_M \eta^2\, \psi^2\, dV}. \tag{93.16}$$

Proof. Equation (93.15) follows from Dirichlet-Neumann bracketing [55, Chapter XIII.15]. To prove (93.16), $\eta\psi$ is supported in $\mathrm{int}(X)$ and so

$$\begin{aligned} \lambda_X &\leq \frac{\int_M \left(4|\nabla(\eta\psi)|^2 + R\,\eta^2\psi^2\right)\, dV}{\int_M \eta^2\psi^2\, dV} \\ &= \frac{\int_M \left(4|\nabla\eta|^2\psi^2 + 8\langle \nabla\eta, \nabla\psi\rangle \eta\psi + 4\eta^2|\nabla\psi|^2 + R\,\eta^2\psi^2\right)\, dV}{\int_M \eta^2\psi^2\, dV}. \end{aligned} \tag{93.17}$$

As $-4\triangle\psi + R\psi = \lambda_M\psi$, we have

$$\lambda_M \int_M \eta^2\psi^2 dV = -4\int_M \eta^2\psi\triangle\psi\, dV + \int_M R\eta^2\psi^2\, dV \tag{93.18}$$
$$= 4\int_M \langle\nabla(\eta^2\psi),\nabla\psi\rangle\, dV + \int_M R\eta^2\psi^2\, dV$$
$$= 4\int_M \eta^2\, |\nabla\psi|^2\, dV + 8\int_M \langle\nabla\eta,\nabla\psi\rangle\eta\psi\, dV + \int_M R\eta^2\psi^2\, dV.$$

Equation (93.16) follows. □

The next result is an Agmon-type estimate.

Lemma 93.19. *With the notation of Lemma 93.14, given a nonnegative function $\phi \in C^\infty(M)$, suppose that $f \in C^\infty(M)$ satisfies*

$$4|\nabla f|^2 \le R - \lambda_M - c \tag{93.20}$$

on $\mathrm{supp}(\phi)$*, for some $c > 0$. Then*

$$\|e^f\phi\psi\|_2 \le 4\, c^{-1}\left(\|e^f\triangle\phi\|_\infty + \|e^f\nabla\phi\|_\infty(\lambda_M - \min R)^{1/2}\right)\|\psi\|_2. \tag{93.21}$$

Proof. Put $H = -4\triangle + R$. By assumption,

$$\phi(R - 4|\nabla f|^2 - \lambda_M)\,\phi \ge c\,\phi^2 \tag{93.22}$$

and so there is an inequality of operators on $L^2(M)$:

$$\phi(H - 4|\nabla f|^2 - \lambda_M)\,\phi = 4\phi d^* d\phi + \phi(R - 4|\nabla f|^2 - \lambda_M)\,\phi \ge c\,\phi^2. \tag{93.23}$$

In particular,

$$\int_M e^f\psi\phi\,(H - 4|\nabla f|^2 - \lambda_M)\,\phi e^f\psi dV \ge c\int_M \phi^2 e^{2f}\psi^2\, dV. \tag{93.24}$$

For $\rho \in C^\infty(M)$,

$$e^f H(e^{-f}\rho) = H\rho + 4\,\nabla\cdot((\nabla f)\rho) + 4\,\langle\nabla f,\nabla\rho\rangle - 4\,|\nabla f|^2\rho \tag{93.25}$$

and so

$$\int_M \rho e^f H(e^{-f}\rho)\, dV = \int_M \rho(H - 4\,|\nabla f|^2)\rho\, dV. \tag{93.26}$$

Taking $\rho = e^f\phi\psi$ gives

$$\int_M e^{2f}\phi\psi H(\phi\psi)\, dV = \int_M e^f\phi\psi(H - 4\,|\nabla f|^2)e^f\phi\psi\, dV. \tag{93.27}$$

From (93.24) and (93.27),
(93.28)
$$c\|e^f\phi\psi\|_2^2 \le \int_M e^{2f}\phi\psi(H - \lambda_M)(\phi\psi)\, dV = \int_M e^{2f}\phi\psi[H,\phi]\psi\, dV = \langle e^f\phi\psi, e^f[H,\phi]\psi\rangle_2.$$

Thus

$$c\|e^f\phi\psi\|_2 \le \|e^f[H,\phi]\psi\|_2. \tag{93.29}$$

Now

$$e^f [H, \phi] \psi \ = \ - 4 \, e^f (\triangle \phi) \psi \ - \ 8 \, e^f \langle \nabla \phi, \nabla \psi \rangle. \tag{93.30}$$

Then

$$\| e^f [H, \phi] \psi \|_2 \ \le \ 4 \, \| e^f \triangle \phi \|_\infty \, \| \psi \|_2 \ + \ 8 \, \| e^f \nabla \phi \|_\infty \, \| \nabla \psi \|_2. \tag{93.31}$$

Finally,

$$4 \, \| \nabla \psi \|_2^2 \ = \ \int_M (\lambda_M \ - \ R) \, \psi^2 \, dV \tag{93.32}$$

and so

$$2 \| \nabla \psi \|_2 \ \le \ (\lambda_M \ - \ \min R)^{1/2} \, \| \psi \|_2. \tag{93.33}$$

This proves the lemma. □

Clearly Lemma 93.19 is also true if f is just assumed to be Lipschitz-regular.

We now apply Lemmas 93.14 and 93.19 to a Ricci flow with surgery. A singularity caused by extinction of a component will not be a problem, so let T_0 be a surgery time and let $M_+ = \mathcal{M}^+_{T_0}$ be the postsurgery manifold. We will write λ^+ instead of λ_{M_+}. Let $M_{cap} = \mathcal{M}^+_{T_0} - (\mathcal{M}^+_{T_0} \cap \mathcal{M}^-_{T_0})$ be the added caps and put $X \ = \ \overline{M_+ - M_{cap}} = \overline{\mathcal{M}^+_{T_0} \cap \mathcal{M}^-_{T_0}}$. For simplicity, let us assume that M_{cap} has a single component; the argument in the general case is similar. From the nature of the surgery procedure, the surgery is done in an ϵ-horn extending from Ω_ρ, where $\rho \ = \ \delta(T_0) r(T_0)$. In fact, because of the canonical neighborhood assumption, we can extend the ϵ-horn inward until $R \sim r(T_0)^{-2}$. Appplying (93.1) with a test function supported in an ϵ-tube near this inner boundary, it follows that $\lambda^+ \ \le \ c' r(T_0)^{-2}$ for some universal constant $c' >> 1$.

In what follows we take $\delta(T_0)$ to be small. As R is much greater than $r(T_0)^{-2}$ on M_{cap}, it follows that $\lambda_{M - \mathrm{int}(X)}$ is much greater than $r(T_0)^{-2}$. Then from (93.15), $\lambda^+ \ \le \ \lambda_X$. We can apply (93.16) to get an inequality the other way. We take the function η to interpolate from being 1 outside of the $h(T_0)$-neighborhood $N_h M_{cap}$ of M_{cap}, to being 0 on M_{cap}. In terms of the normalized eigenfunction ψ on M_+, this gives a bound of the form

$$\lambda_X \ \le \ \lambda^+ \ + \ \mathrm{const.} \ h(T_0)^{-2} \, \frac{\int_{N_h M_{cap}} \psi^2 \, dV}{1 - \int_{N_h M_{cap}} \psi^2 \, dV}. \tag{93.34}$$

We now wish to show that $\int_{N_h M_{cap}} \psi^2 \, dV$ is small. For this we apply Lemma 93.19 with $c \ = \ c' \, r(T_0)^{-2}$. Take an ϵ-tube U, in the ϵ-horn, whose center has scalar curvature roughly $200 \, c' \, r(T_0)^{-2}$ and which is the closest tube to the cap with this property. Let $x \ : \ U \ \rightarrow \ (- \, \epsilon^{-1}, \epsilon^{-1})$ be the longitudinal parametrization of the tube, which we take to be increasing in the direction of the surgery cap. Let $\Phi \ : \ (-1, 1) \ \rightarrow \ [0, 1]$ be a fixed nondecreasing smooth function which is zero on $(-1, 1/4)$ and one on $(1/2, 1)$. Put $\phi = \Phi \circ x$ on U. Extend ϕ to M_+ by making it zero to the left of U and one to the right of U, where "right of U" means the connected component of $M_+ - U$ containing the surgery cap. Dimensionally, $|\nabla \phi|_\infty \ \le \ \mathrm{const.} \ r(T_0)^{-1}$ and $|\triangle \phi|_\infty \ \le \ \mathrm{const.} \ r(T_0)^{-2}$. Define a function f to the right of $x^{-1}(0)$ by setting it to be the distance from $x^{-1}(0)$ with respect to the metric $\frac{1}{4} \, (R - \lambda^+ - c) \, g_{M_+}$. (Note that to the right of $x^{-1}(0)$, we have $R \ \ge \ 200 c' r(T_0)^{-2} \ \ge$

$\lambda^+ + c$.) Then equation (93.21) gives $|e^f \phi \psi|_2 \le$ const.(T_0). The point is that const.(T_0) is independent of the (small) surgery parameter $\delta(T_0)$.

Hence

$$\int_{N_h M_{cap}} \psi^2 \, dV \le \left(\sup_{N_h M_{cap}} e^{-2f} \right) \int_{N_h M_{cap}} e^{2f} \psi^2 \, dV \le \text{const.} \sup_{N_h M_{cap}} e^{-2f}. \tag{93.35}$$

To estimate $\sup_{N_h M_{cap}} e^{-2f}$, we use the fact that the ϵ-horn consists of a sequence of ϵ-tubes stacked together. In the region of M_+ from $x^{-1}(0)$ to the surgery cap, the scalar curvature ranges from roughly $200\, c'\, r(T_0)^{-2}$ to $h(T_0)^{-2}$. On a given ϵ-tube, if ϵ is sufficiently small then the ratio of the scalar curvatures between the two ends is bounded by e. Hence in going from $x^{-1}(0)$ to the surgery cap, one must cross at least N disjoint ϵ-tubes, with $e^N = \frac{1}{200c'}\, r(T_0)^2\, h(T_0)^{-2}$. Traversing a given ϵ-tube (say of radius r') in going towards the surgery cap, f increases by roughly const. $\int_{-\epsilon^{-1} r'}^{\epsilon^{-1} r'} (r')^{-1} \, ds$, which is const. ϵ^{-1}. Hence near the surgery cap, we have

$$\sup_{N_h M_{cap}} e^{-2f} \le \text{const.}\ e^{-\text{const.}\, N \epsilon^{-1}} = \text{const.}\ (r(T_0)^2\, h(T_0)^{-2})^{-\text{const.}\, \epsilon^{-1}}. \tag{93.36}$$

Combining this with (93.34) and (93.35), we obtain

$$\lambda_X \le \lambda^+ + \text{const.}\ h(T_0)^{-2}\, (r(T_0)^2\, h(T_0)^{-2})^{-\text{const.}\, \epsilon^{-1}}. \tag{93.37}$$

By making a single redefinition of ϵ, we can ensure that $\lambda_X \le \lambda^+ +$ const. $h(T_0)^4$. The last constant will depend on $r(T_0)$ but is independent of $\delta(T_0)$. Thus if $\delta(T_0)$ is small enough, we can ensure that $|\lambda_X - \lambda^+|$ is small in comparison to the volume change $V^-(T_0) - V^+(T_0)$, which is comparable to $h(T_0)^3$.

If λ^- is the smallest eigenvalue of $-4\triangle + R$ on the presurgery manifold $\mathcal{M}_t$, for t slightly less than T_0, then we can estimate $|\lambda_X - \lambda^-|$ in a similar way. Hence for an arbitrary positive continuous function $\xi(t)$, we can make the parameters $\overline{\delta}_j$ of Proposition 77.2 small enough to ensure that

$$|\lambda^+(T_0) - \lambda^-(T_0)| \le \xi(T_0)\, (V^-(T_0) - V^+(T_0)) \tag{93.38}$$

for a surgery at time T_0.

We now redo the argument for the proposition, as given above in the surgery-free case, in the presence of surgeries. Suppose first that $\lambda(g_0) > 0$ for some metric g_0 on M. After possible rescaling, we can assume that g_0 is the initial condition for a Ricci flow with surgery $(\mathcal{M}, g(\cdot))$, with normalized initial condition. Using the lower scalar curvature bound of Lemma 79.11 and the Ricci flow equation, the volume on the time interval $[0, \frac{3}{\lambda(0)}]$ has an *a priori* upper bound of the form const. $V(0)$. As a surgery at time T_0 removes a volume comparable to $h(T_0)^3$, we have $\sum h(T_0)^3 \le$ const. $V(0)$, where the sum is over the surgeries and T_0 denotes the surgery time. From the above discussion, the change in λ due to the surgeries is bounded below by $-$ const. $\sum_{T_0} h(T_0)^4$. Then the decrease in λ due to surgeries

on the time interval $[0, \frac{3}{\lambda(0)}]$ is bounded above by

$$\text{const.} \sum_{T_0 \in [0, \frac{3}{\lambda(0)}]} h(T_0)^4 \, \leq \, \text{const.} \left(\sup_{t \in [0, \frac{3}{\lambda(0)}]} h(t) \right) V(0). \tag{93.39}$$

By choosing the function $\delta(t)$ to be sufficiently small, the decrease in λ due to surgeries is not enough to prevent the blowup of λ on the time interval $[0, \frac{3}{\lambda(0)}]$ coming from the increase of λ between the surgeries. Hence the solution goes extinct.

Now suppose that M does not admit a metric g with $\lambda(g) > 0$. Again, if M is a graph manifold then $\overline{\lambda} \, = \, 0$.

Suppose that M is not a graph manifold. As before, $\overline{\lambda} \, \geq \, - \frac{3}{2} \, \widehat{V}^{2/3}$. Given an initial metric g_0, we wish to show that by choosing the function $\delta(t)$ small enough we can make the function $\lambda(t) V^{2/3}(t)$ arbitrarily close to being nondecreasing. To see this, we consider the effect of a surgery on $\lambda(t) V^{2/3}(t)$. Upon performing a surgery the volume decreases, which in itself cannot decrease $\lambda(t) V^{2/3}(t)$. (We are using the fact that $\lambda(t)$ is nonpositive.) Then from the above discussion, the change in $\lambda(t) V^{2/3}(t)$ due from a surgery at time T_0, is bounded below by $-\xi(T_0) \, (V^-(T_0) - V^+(T_0)) \, V^-(T_0)^{2/3}$. With normalized initial conditions, we have an *a priori* upper bound on $V(t)$ in terms of $V(0)$ and t. Over any time interval $[T_1, T_2]$, we must have

$$\sum_{T_0 \in [T_1, T_2]} \big(V^-(T_0) - V^+(T_0) \big) \, \leq \, \sup_{t \in [T_1, T_2]} V(t), \tag{93.40}$$

where the sum is over the surgeries in the interval $[T_1, T_2]$. Then along with the monotonicity of $\lambda(t) V^{2/3}(t)$ in between the surgery times, by choosing the function $\delta(t)$ appropriately we can ensure that for any $\sigma > 0$ there is a Ricci flow with (r, δ)-cutoff starting from g_0 so that $\lambda(g_0) \, V^{2/3}(g_0) \, \leq \, \lambda(t) \, V^{2/3}(t) \, + \, \sigma$ for all t. It follows that $\overline{\lambda} \, \leq \, - \frac{3}{2} \, \widehat{V}^{2/3}$.

This shows that $\overline{\lambda} \, = \, - \frac{3}{2} \, \widehat{V}^{2/3}$. The same argument as before shows that if we have a decomposition with the hyperbolic volume of N equal to $\widehat{V}$ then the cusps of N (if any) are incompressible in M'. □

Remark 93.41. It follows that if the three-manifold M does not admit a metric of positive scalar curvature then $\sigma(M) \, = \, \overline{\lambda}$. In fact, this is true in any dimension $n \geq 3$ [1].

Appendix A. Maximum principles

In this appendix we list some maximum principles and their consequences. Our main source is [23], where references to the original literature can be found.

The first type of maximum principle is a *weak* maximum principle which says that under certain conditions, a spatial inequality on the initial condition implies a time-dependent inequality at later times.

Theorem A.1. *Let M be a closed manifold. Let $\{g(t)\}_{t \in [0,T]}$ be a smooth one-parameter family of Riemannian metrics on M and let $\{X(t)\}_{t \in [0,T]}$ be a smooth one-parameter family*

of vector fields on M. Let $F : \mathbb{R} \times [0,T] \rightarrow \mathbb{R}$ be a Lipschitz function. Suppose that $u = u(x,t)$ is C^2-regular in x, C^1-regular in t and

$$\frac{\partial u}{\partial t} \leq \triangle_{g(t)} u + X(t) u + F(u,t). \tag{A.2}$$

Let $\phi : [0,T] \rightarrow \mathbb{R}$ be the solution of $\frac{d\phi}{dt} = F(\phi(t),t)$ with $\phi(0) = \alpha$. If $u(\cdot, 0) \leq \alpha$ then $u(\cdot,t) \leq \phi(t)$ for all $t \in [0,T]$.

There are various noncompact versions of the weak maximum principle. We state one here.

Theorem A.3. *Let $(M, g(\cdot))$ be a complete Ricci flow solution on the interval $[0,T]$ with uniformly bounded curvature. If $u = u(x,t)$ is a $W^{1,2}_{loc}$ function that weakly satisfies $\frac{\partial u}{\partial t} \leq \triangle_{g(t)} u$, with $u(\cdot, 0) \leq 0$ and*

$$\int_0^T \int_M e^{- c\, d_t^2(x,x_0)}\, u^2(x,s)\, dV(x)\, ds < \infty \tag{A.4}$$

for some $c > 0$, then $u(\cdot,t) \leq 0$ for all $t \in [0,T]$.

A *strong* maximum principle says that under certain conditions, a strict inequality at a given time implies strict inequality at later times and also slightly earlier times. It does not require complete metrics.

Theorem A.5. *Let M be a connected manifold. Let $\{g(t)\}_{t \in [0,T]}$ be a smooth one-parameter family of Riemannian metrics on M and let $\{X(t)\}_{t \in [0,T]}$ be a smooth one-parameter family of vector fields on M. Let $F : \mathbb{R} \times [0,T] \rightarrow \mathbb{R}$ be a Lipschitz function. Suppose that $u = u(x,t)$ is C^2-regular in x, C^1-regular in t and*

$$\frac{\partial u}{\partial t} \leq \triangle_{g(t)} u + X(t) u + F(u,t). \tag{A.6}$$

Let $\phi : [0,T] \rightarrow \mathbb{R}$ be a solution of $\frac{d\phi}{dt} = F(\phi(t),t)$. If $u(\cdot,t) \leq \phi(t)$ for all $t \in [0,T]$ and $u(x_0,t_0) < \phi(t)$ for some $x_0 \in M$ and $t_0 \in (0,T]$ then there is some $\epsilon > 0$ so that $u(\cdot,t) < \phi(t)$ for $t \in (t_0 - \epsilon, T]$.

A consequence of the strong maximum principle is a statement about restricted holonomy for Ricci flow solutions with nonnegative curvature operator Rm.

Theorem A.7. *Let M be a connected manifold. Let $\{g(t)\}_{t \in [0,T]}$ be a smooth one-parameter family of Riemannian metrics on M with nonnegative curvature operator that satisfy the Ricci flow equation. Then for each $t \in (0,T]$, the image $\mathrm{Im}(\mathrm{Rm}_{g(t)})$ of the curvature operator is a smooth subbundle of $\Lambda^2(T^*M)$ which is invariant under spatial parallel translation. There is a sequence of times $0 = t_0 < t_1 < \ldots < t_k = T$ such that for each $1 \leq i \leq k$, $\mathrm{Im}(\mathrm{Rm}_{g(t)})$ is a Lie subalgebra of $\Lambda^2(T^*_m M) \cong o(n)$ that is independent of t for $t \in (t_{i-1}, t_i]$. Furthermore, $\mathrm{Im}(\mathrm{Rm}_{g(t_i)}) \subset \mathrm{Im}(\mathrm{Rm}_{g(t_{i+1})})$.*

In particular, under the hypotheses of Theorem A.7, a local isometric splitting at a given time implies a local isometric splitting at earlier times.

Appendix B. ϕ-almost nonnegative curvature

In three dimensions, the Ricci flow equation implies that

$$\frac{dR}{dt} = \triangle R + \frac{2}{3}R^2 + 2\,|R_{ij} - \frac{R}{3}g_{ij}|^2. \tag{B.1}$$

The maximum principle of Appendix A implies that if $(M, g(\cdot))$ is a Ricci flow solution defined for $t \in [0, T)$, with complete time slices and bounded curvature on compact time intervals, then

$$(\inf R)(t) \geq \frac{(\inf R)(0)}{1 - \frac{2}{3}t(\inf R)(0)}. \tag{B.2}$$

In particular, $tR(\cdot, t) > -\frac{3}{2}$ for all $t \geq 0$ (compare [34, Section 2]).

Recall that the curvature operator is an operator on 2-forms. We follow the usual Ricci flow convention that if a manifold has constant sectional curvature k then its curvature operator is multiplication by $2k$. In general, the trace of the curvature operator equals the scalar curvature.

In three dimensions, having nonnegative curvature operator is equivalent to having nonnegative sectional curvature. Each eigenvalue of the curvature operator is twice a sectional curvature.

Hamilton-Ivey pinching, as given in [34, Theorem 4.1], says the following.

Assume that at $t = 0$ the eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ of the curvature operator at each point satisfy $\lambda_1 \geq -1$. (One can always achieve this by rescaling. Note that it implies $R(\cdot, t) \geq -\frac{3}{2}\frac{1}{t+\frac{1}{4}}$.) Given a point (x, t), put $X = -\lambda_1$. If $X > 0$ then

$$R(x, t) \geq X\left(\ln X + \ln(1 + t) - 3\right), \tag{B.3}$$

or equivalently,

$$tR(x, t) \geq tX\left(\ln(tX) + \ln(\frac{1+t}{t}) - 3\right). \tag{B.4}$$

Definition B.5. Given $t \geq 0$, a Riemannian 3-manifold (M, g) satisfies the *time-t Hamilton-Ivey pinching condition* if for every $x \in M$, if $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the eigenvalues of the curvature operator at x, then either

- $\lambda_1 \geq 0$, i.e. the curvature is nonnegative at x,

or

- If $\lambda_1 < 0$ and $X = -\lambda_1$ then $tR(x) \geq tX\left(\ln(tX) + \ln\left(\frac{1+t}{t}\right) - 3\right)$.

This condition has the following monotonicity property:

Lemma B.6. *Suppose that* Rm *and* Rm$'$ *are 3-dimensional curvature operators whose scalar curvatures and first eigenvalues satisfy* $R' \geq R \geq 0$ *and* $\lambda_1' \geq \lambda_1$. *If* Rm *satisfies the time-t Hamilton-Ivey pinching condition then so does* Rm$'$.

Proof. We may assume that $\lambda_1' < 0$ and $\log(tX') + \ln\left(\frac{1+t}{t}\right) - 3 > 0$, since otherwise the condition will be satisfied (because $R' > 0$ by hypothesis). The function

$$Y \mapsto tY\left(\ln(tY) + \ln\left(\frac{1+t}{t}\right) - 3\right) \tag{B.7}$$

is monotone increasing on the interval on which $\ln(tY) + \ln\left(\frac{1+t}{t}\right) - 3$ is nonnegative, so $tX\left(\ln(tX) + \ln\left(\frac{1+t}{t}\right) - 3\right) \geq tX'\left(\ln(tX') + \ln\left(\frac{1+t}{t}\right) - 3\right)$. Hence Rm' satisfies the pinching condition too. ☐

The content of the pinching equation is that for any $s \in \mathbb{R}$, if $tR(\cdot,t) \leq s$ then there is a lower bound $t\,\mathrm{Rm}(\cdot,t) \geq \mathrm{const.}(s,t)$. Of course, this is a vacuous statement if $s \leq -\frac{3}{2}$.

Using equation (B.4), we can find a positive function $\Phi \in C^\infty(\mathbb{R})$ such that

1. Φ is nondecreasing.
2. For $s > 0$, $\frac{\Phi(s)}{s}$ is decreasing.
3. For large s, $\Phi(s) \sim \frac{s}{\ln s}$.
4. For all t,

$$\mathrm{Rm}(\cdot,t) \geq -\Phi(R(\cdot,t)). \tag{B.8}$$

This bound has the most consequence when s is large.

We note that for the original unscaled Ricci flow solution, the precise bound that we obtain depends on t_0 and the time-zero metric, through its lower curvature bound.

Appendix C. Ricci solitons

Let $\{V(t)\}$ be a time-dependent family of vector fields on a manifold M. The solution to the equation

$$\frac{dg}{dt} = \mathcal{L}_{V(t)}g \tag{C.1}$$

is

$$g(t) = \phi^{-1}(t)^* g(t_0) \tag{C.2}$$

where $\{\phi(t)\}$ is the 1-parameter group of diffeomorphisms generated by $-V$, normalized by $\phi(t_0) = \mathrm{Id}$. (If M is noncompact then we assume that V can be integrated. The reason for the funny signs is that if a 1-parameter family of diffeomorphisms $\eta(t)$ is generated by vector fields $W(t)$ then $\mathcal{L}_{W(t)} = \eta^{-1}(t)^* \left.\frac{d\eta(t+\epsilon)^*}{d\epsilon}\right|_{\epsilon=0} = -\left.\frac{d\eta^{-1}(t+\epsilon)^*}{d\epsilon}\right|_{\epsilon=0}\eta(t)^*$.)

The equation for a *steady soliton* is

$$2\,\mathrm{Ric} + \mathcal{L}_V g = 0, \tag{C.3}$$

where V is a time-independent vector field. The corresponding Ricci flow is given by

$$g(t) = \phi^{-1}(t)^* g(t_0), \tag{C.4}$$

where $\{\phi(t)\}$ is the 1-parameter group of diffeomorphisms generated by $-V$. (Of course, in this case $\{\phi^{-1}(t)\}$ is the 1-parameter group of diffeomorphisms generated by V.)

A *gradient steady soliton* satisfies the equations

$$\text{(C.5)} \qquad \begin{aligned} \frac{\partial g_{ij}}{\partial t} &= -\, 2R_{ij} \;=\; 2\nabla_i \nabla_j f, \\ \frac{\partial f}{\partial t} &= |\nabla f|^2. \end{aligned}$$

It follows from (C.5) that

$$\text{(C.6)} \qquad \frac{\partial}{\partial t}\left(g^{ij}\,\partial_j f\right) \;=\; 2\,R^{ij}\,\nabla_j f \;+\; g^{ij}\nabla_j|\nabla f|^2 \;=\; -2\,(\nabla^i\nabla^j f)\,\nabla_j f \;+\; \nabla^i|\nabla f|^2 \;=\; 0,$$

showing that $V \;=\; \nabla f$ is indeed constant in t. The solution to (C.5) is

$$\text{(C.7)} \qquad \begin{aligned} g(t) &= \phi^{-1}(t)^* g(t_0), \\ f(t) &= \phi^{-1}(t)^* f(t_0). \end{aligned}$$

Conversely, given a metric $\widehat{g}$ and a function $\widehat{f}$ satisfying

$$\text{(C.8)} \qquad \widehat{R}_{ij} \;+\; \widehat{\nabla}_i\widehat{\nabla}_j\widehat{f} \;=\; 0,$$

put $V \;=\; \widehat{\nabla}\widehat{f}$. If we define $g(t)$ and $f(t)$ by

$$\text{(C.9)} \qquad \begin{aligned} g(t) &= \phi^{-1}(t)^*\widehat{g}, \\ f(t) &= \phi^{-1}(t)^*\widehat{f} \end{aligned}$$

then they satisfy (C.5).

A solution to (C.5) satisfies

$$\text{(C.10)} \qquad \frac{\partial f}{\partial t} \;=\; |\nabla f|^2 \;-\; \triangle f \;-\; R,$$

or

$$\text{(C.11)} \qquad \frac{\partial}{\partial t}\,e^{-f} \;=\; -\,\triangle e^{-f} \;+\; R\,e^{-f}.$$

This perhaps motivates Perelman's use of the backward heat equation (5.23).

A *shrinking soliton* lives on a time interval $(-\infty, T)$. For convenience, we take $T \;=\; 0$. Then the equation is

$$\text{(C.12)} \qquad 2\,\mathrm{Ric} + \mathcal{L}_V g \;+\; \frac{g}{t} \;=\; 0.$$

The vector field $V = V(t)$ satisfies $V(t) \;=\; -\,\frac{1}{t}\,V(-1)$. The corresponding Ricci flow is given by

$$\text{(C.13)} \qquad g(t) \;=\; -\,t\,\phi^{-1}(t)^* g(-1),$$

where $\{\phi(t)\}$ is the 1-parameter group of diffeomorphisms generated by $-V$, normalized by $\phi(-1) \;=\; \mathrm{Id}$.

A *gradient shrinking soliton* satisfies the equations

$$\frac{\partial g_{ij}}{\partial t} = -2R_{ij} = 2\nabla_i\nabla_j f + \frac{g_{ij}}{t}, \qquad \text{(C.14)}$$
$$\frac{\partial f}{\partial t} = |\nabla f|^2.$$

It follows from (C.14) that $V = \nabla f$ satisfies $V(t) = -\frac{1}{t}V(-1)$. The solution to (C.14) is

$$g(t) = -t\,\phi^{-1}(t)^* g(-1), \qquad \text{(C.15)}$$
$$f(t) = \phi^{-1}(t)^* f(-1).$$

Conversely, given a metric $\widehat{g}$ and a function $\widehat{f}$ satisfying

$$\widehat{R}_{ij} + \widehat{\nabla}_i\widehat{\nabla}_j\widehat{f} - \frac{1}{2}\widehat{g} = 0, \qquad \text{(C.16)}$$

put $V(t) = -\frac{1}{t}\widehat{\nabla}\widehat{f}$. If we define $g(t)$ and $f(t)$ by

$$g(t) = -t\,\phi^{-1}(t)^*\widehat{g}, \qquad \text{(C.17)}$$
$$f(t) = \phi^{-1}(t)^*\widehat{f}$$

then they satisfy (C.14).

An *expanding soliton* lives on a time interval (T, ∞). For convenience, we take $T = 0$. Then the equation is

$$2\,\mathrm{Ric} + \mathcal{L}_V g + \frac{g}{t} = 0. \qquad \text{(C.18)}$$

The vector field $V = V(t)$ satisfies $V(t) = \frac{1}{t}V(1)$. The corresponding Ricci flow is given by

$$g(t) = t\,\phi^{-1}(t)^* g(1), \qquad \text{(C.19)}$$

where $\{\phi(t)\}$ is the 1-parameter group of diffeomorphisms generated by $-V$, normalized by $\phi(1) = \mathrm{Id}$.

A *gradient expanding soliton* satisfies the equations

$$\frac{\partial g_{ij}}{\partial t} = -2R_{ij} = 2\nabla_i\nabla_j f + \frac{g_{ij}}{t}, \qquad \text{(C.20)}$$
$$\frac{\partial f}{\partial t} = |\nabla f|^2.$$

It follows from (C.20) that $V = \nabla f$ satisfies $V(t) = \frac{1}{t}V(1)$. The solution to (C.20) is

$$g(t) = t\,\phi^{-1}(t)^* g(1), \qquad \text{(C.21)}$$
$$f(t) = \phi^{-1}(t)^* f(1).$$

Conversely, given a metric $\widehat{g}$ and a function $\widehat{f}$ satisfying

$$\widehat{R}_{ij} + \widehat{\nabla}_i\widehat{\nabla}_j\widehat{f} + \frac{1}{2}\widehat{g} = 0, \qquad \text{(C.22)}$$

put $V(t) = \frac{1}{t} \widehat{\nabla} \widehat{f}$. If we define $g(t)$ and $f(t)$ by

$$\begin{aligned} g(t) &= t\, \phi^{-1}(t)^* \widehat{g}, \\ f(t) &= \phi^{-1}(t)^* \widehat{f} \end{aligned} \tag{C.23}$$

then they satisfy (C.20).

Obvious examples of solitons are given by Einstein metrics, with $V = 0$. Any steady or expanding soliton on a closed manifold comes from an Einstein metric. Other examples of solitons (see [22, Chapter 2]) are :

1. (Gradient steady soliton) The cigar soliton on $\mathbb{R}^2$ and the Bryant soliton on $\mathbb{R}^3$.
2. (Gradient shrinking soliton) Flat $\mathbb{R}^n$ with $f = -\frac{|x|^2}{4t}$.
3. (Gradient shrinking soliton) The shrinking cylinder $\mathbb{R} \times S^{n-1}$ with $f = -\frac{x^2}{4t}$, where x is the coordinate on $\mathbb{R}$.
4. (Gradient shrinking soliton) The Koiso soliton on $\mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$.

Appendix D. Local derivative estimates

Theorem D.1. *For any $\alpha, K, K', l \geq 0$ and $m, n \in \mathbb{Z}^+$, there is some $C = C(\alpha, K, K', l, m, n)$ with the following property. Given $r > 0$, suppose that $g(t)$ is a Ricci flow solution for $t \in [0, \overline{t}]$, where $0 < \overline{t} \leq \frac{\alpha r^2}{K}$, defined on an open neighborhood U of a point $p \in M^n$. Suppose that $\overline{B(p, r, 0)}$ is a compact subset of U, that*

$$| \operatorname{Rm}(x,t)| \leq \frac{K}{r^2} \tag{D.2}$$

for all $x \in U$ and $t \in [0, \overline{t}]$, and that

$$|\nabla^\beta \operatorname{Rm}(x,0)| \leq \frac{K'}{r^{|\beta|+2}} \tag{D.3}$$

for all $x \in U$ and $|\beta| \leq l$. Then

$$|\nabla^\beta \operatorname{Rm}(x,t)| \leq \frac{C}{r^{|\beta|+2} \left(\frac{t}{r^2}\right)^{\frac{\max(m-l,0)}{2}}} \tag{D.4}$$

for all $x \in B\left(p, \frac{r}{2}, 0\right)$, $t \in (0, \overline{t}]$ and $|\beta| \leq m$.

In particular, $|\nabla^\beta \operatorname{Rm}(x,t)| \leq \frac{C}{r^{|\beta|+2}}$ whenever $|\beta| \leq l$.

The main case $l = 0$ of Theorem D.1 is due to Shi [62]. The extension to $l \geq 0$ appears in [41, Appendix B].

Appendix E. Convergent subsequences of Ricci flow solutions

Theorem E.1. [32] *Let $\{g_i(t)\}_{i=1}^\infty$ be a sequence of Ricci flow solutions on connected pointed manifolds (M_i, m_i), defined for $t \in (A, B)$ and complete for each i and t, with $-\infty \leq A < 0 < B \leq \infty$. Suppose that the following two conditions are satisfied :*

1. For each $r > 0$ and each compact interval $I \subset (A, B)$, there is an $N_{r,I} < \infty$ so that for

all $t \in I$ and all i, $\sup_{B_t(m_i,r)} |\operatorname{Rm}(g_i(t))| \le N_{r,I}$, and
2. The time-0 injectivity radii $\{\operatorname{inj}(g_i(0))(m_i)\}_{i=1}^{\infty}$ are uniformly bounded below by a positive number.

Then after passing to a subsequence, the solutions converge smoothly to a complete Ricci flow solution $g_\infty(t)$ on a connected pointed manifold (M_∞, m_∞), defined for $t \in (A,B)$. That is, for any compact interval $I \subset (A,B)$ and any compact set $K \subset M_\infty$ containing m_∞, there are pointed time-independent diffeomorphisms $\phi_{K,i} : K \to K_i$ (with $K_i \subset M_i$) so that $\{\phi_{K,i}^* g_i\}_{i=1}^{\infty}$ converges smoothly to g_∞ on $I \times K$.

Given the sectional curvature bounds, the lower bound on the injectivity radii is equivalent to a lower bound on the volumes of balls around m_0 [20, Theorem 4.7].

There are many variants of the theorem with alternative hypotheses. One can replace the interval (A,B) with an interval $(A,B]$, $-\infty \le A < 0 \le B < \infty$. One can also replace the interval (A,B) with an interval $[A,B)$, $-\infty < A < 0 < B \le \infty$, if in addition one has uniform time-A bounds $\sup_{B_A(m_i,r)} |\nabla^j \operatorname{Rm}(g_i(A))| \le C_{r,j}$. Then using Appendix D, one gets smooth convergence to a limit solution g_∞ on the time interval $[A,B)$. (Without the time-A bounds one would only get C^0-convergence on $[A,B)$ and C^∞-convergence on (A,B).) There is a similar statement if one replaces the interval (A,B) with an interval $[A,B]$, $-\infty < A < 0 \le B < \infty$.

In the incomplete setting, if one has curvature bounds on spacetime product regions $I \times B_0(m_i,r)$, with $I \subset (A,B)$ and $r < r_0 < \infty$, and an injectivity radius bound at $(m_i,0)$, then one can still extract a subsequence which converges on compact subsets of (A,B) times compact subsets of $B_0(m_\infty, r_0)$. There are analogous statements when (A,B) is replaced by $(A,B]$, $[A,B)$ or $[A,B]$, and the ball is replaced by a metric annulus.

Appendix F. Harnack inequalities for Ricci flow

We first recall the statement of the matrix Harnack inequality. Put

$$\text{(F.1)} \qquad \begin{aligned} P_{abc} &= \nabla_a R_{bc} - \nabla_b R_{ac}, \\ M_{ab} &= \triangle R_{ab} - \frac{1}{2} \nabla_a \nabla_b R + 2\, R_{acbd} R_{cd} - R_{ac} R_{bc} + \frac{R_{ab}}{2t}. \end{aligned}$$

Given a 2-form U and a 1-form W, put

$$\text{(F.2)} \qquad Z(U,W) = M_{ab} W_a W_b + 2\, P_{abc} U_{ab} W_c + R_{abcd} U_{ab} U_{cd}.$$

Suppose that we have a Ricci flow for $t > 0$ on a complete manifold with bounded curvature on each compact time interval and nonnegative curvature operator. Hamilton's matrix Harnack inequality says that for all $t > 0$ and all U and W, $Z(U,W) \ge 0$ [33, Theorem 14.1].

Taking $W_a = Y_a$ and $U_{ab} = (X_a Y_b - Y_a X_b)/2$ and using the fact that

$$\text{(F.3)} \qquad \operatorname{Ric}_t(Y,Y) = (\triangle R_{ab}) Y^a Y^b + 2\, R_{acbd} R_{cd} Y^a Y^b - 2\, R_{ac} R_{bc} Y^a Y^b,$$

we can write $2Z(U,W) = H(X,Y)$, where

(F.4)
$$H(X,Y) = -\operatorname{Hess}_R(Y,Y) - 2\langle R(Y,X)Y,X\rangle + 4\left(\nabla_X \operatorname{Ric}(Y,Y) - \nabla_Y \operatorname{Ric}(Y,X)\right) + 2\operatorname{Ric}_t(Y,Y) + 2\left|\operatorname{Ric}(Y,\cdot)\right|^2 + \frac{1}{t}\operatorname{Ric}(Y,Y).$$

Substituting the elements of an orthonormal basis $\{e_i\}_{i=1}^n$ for Y and summing over i gives

(F.5)
$$\sum_i H(X,e_i) = -\triangle R + 2\operatorname{Ric}(X,X) + 4(\langle\nabla R, X\rangle - \sum_i \nabla_{e_i}\operatorname{Ric}(e_i,X)) + 2\sum_i \operatorname{Ric}_t(e_i,e_i) + 2|\operatorname{Ric}|^2 + \frac{1}{t}R.$$

Tracing the second Bianchi identity gives

$$\sum_i \nabla_{e_i}\operatorname{Ric}(e_i,X) = \frac{1}{2}\langle\nabla R, X\rangle. \tag{F.6}$$

From (F.3),

$$\sum_i \operatorname{Ric}_t(e_i,e_i) = \triangle R. \tag{F.7}$$

Putting this together gives

$$\sum_i H(X,e_i) = H(X), \tag{F.8}$$

where

$$H(X) = R_t + \frac{1}{t}R + 2\langle\nabla R, X\rangle + 2\operatorname{Ric}(X,X). \tag{F.9}$$

We obtain Hamilton's trace Harnack inequality, saying that $H(X) \geq 0$ for all X.

In the rest of this section we assume that the solution is defined for all $t \in (-\infty, 0)$. Changing the origin point of time, we have

$$R_t + \frac{1}{t-t_0}R + 2\langle\nabla R, X\rangle + 2\operatorname{Ric}(X,X) \geq 0 \tag{F.10}$$

whenever $t_0 \leq t$. Taking $t_0 \to -\infty$ gives

$$R_t + 2\langle\nabla R, X\rangle + 2\operatorname{Ric}(X,X) \geq 0 \tag{F.11}$$

In particular, taking $X = 0$ shows that the scalar curvature is nondecreasing in t for any ancient solution with nonnegative curvature operator, assuming again that the metric is complete on each time slice with bounded curvature on each compact time interval. More generally,

$$0 \leq R_t + 2\langle\nabla R, X\rangle + 2\operatorname{Ric}(X,X) \leq R_t + 2\langle\nabla R, X\rangle + 2R\langle X,X\rangle. \tag{F.12}$$

If $\gamma : [t_1, t_2] \to M$ is a curve parametrized by s then taking $X = \frac{1}{2} \frac{d\gamma}{ds}$ gives

$$\frac{dR(\gamma(s), s)}{ds} = R_t(\gamma(s), s) + \left\langle \frac{d\gamma}{ds}, \nabla R \right\rangle \geq - \frac{1}{2} R \left\langle \frac{d\gamma}{ds}, \frac{d\gamma}{ds} \right\rangle . \tag{F.13}$$

Integrating $\frac{d \ln R(\gamma(s),s)}{ds}$ with respect to s and using the fact that $g(t)$ is nonincreasing in t gives

$$R(x_2, t_2) \geq \exp \left(- \frac{d^2_{t_1}(x_1, x_2)}{2(t_2 - t_1)} \right) R(x_1, t_1). \tag{F.14}$$

whenever $t_1 < t_2$ and $x_1, x_2 \in M$. (If $n = 2$ then one can replace $\frac{d^2_{t_1}(x_1,x_2)}{2(t_2-t_1)}$ by $\frac{d^2_{t_1}(x_1,x_2)}{4(t_2-t_1)}$.) In particular, if $R(x_2, t_2) = 0$ for some (x_2, t_2) then $g(t)$ must be flat for all t.

Appendix G. Alexandrov spaces

We recall some facts about Alexandrov spaces (see [11, Chapter 10], [12]). Given points p, x, y in a nonnegatively curved Alexandrov space, we let $\widetilde{\angle}_p(x, y)$ denote the comparison angle at p, i.e. the angle of the Euclidean comparison triangle at the vertex corresponding to p.

The Toponogov splitting theorem says that if X is a proper nonnegatively curved Alexandrov space which contains a line, then X splits isometrically as a product $X = \mathbb{R} \times Y$, where Y is a proper, nonnegatively curved Alexandrov space [11, Theorem 10.5.1].

Let M be an n-dimensional Alexandrov space with nonnegative curvature, $p \in M$, and $\lambda_k \to 0$. Then the sequence $(\lambda_k M, p)$ of pointed spaces Gromov-Hausdorff converges to the Tits cone $C_T M$ (the Euclidean cone over the Tits boundary $\partial_T M$) which is a nonnegatively curved Alexandrov space of dimension $\leq n$ [7, p. 58-59]. If the Tits cone splits isometrically as a product $\mathbb{R}^k \times Y$, then M itself splits off a factor of $\mathbb{R}^k$; using triangle comparison, one finds k orthogonal lines passing through a basepoint, and applies the Toponogov splitting theorem.

Now suppose that $x_k \in M$ is a sequence with $d(x_k, p) \to \infty$ and $r_k \in \mathbb{R}^+$ is a sequence with $\frac{r_k}{d(x_k,p)} \to 0$. Then the sequence $(\frac{1}{r_k} M, x_k)$ subconverges to a pointed Alexandrov space (N_∞, x_∞) which splits off a line. To see this, observe that since $(\frac{1}{d(x_k,p)} M, p)$ converges to a cone, we can find a sequence $y_k \in M$ such that $\frac{d(y_k,x_k)}{d(x_k,p)} \to 1$, and $\widetilde{\angle}_{x_k}(p, y_k) \to \pi$. Observe that for any $\rho < \infty$, we can find sequences $p_k \in \overline{px_k}$, $z_k \in \overline{x_k y_k}$ such that $\frac{d(x_k,p_k)}{r_k} \to \rho$, $\frac{d(x_k,z_k)}{r_k} \to \rho$, and by monotonicity of comparison angles [11, Chapter 4.3] we will have $\widetilde{\angle}_{x_k}(p_k, z_k) \to \pi$. Passing to the Gromov-Hausdorff limit, we find $p_\infty, z_\infty \in N_\infty$ such that $d(p_\infty, x_\infty) = d(z_\infty, x_\infty) = \rho$ and $\widetilde{\angle}_{x_\infty}(p_\infty, z_\infty) = \pi$. Since this construction applies for all ρ, it follows that N_∞ contains a line passing through x_∞. Hence, by the Toponogov splitting theorem, it is isometric to a metric product $\mathbb{R} \times N'$ for some Alexandrov space N'.

If M is a complete Riemannian manifold of nonnegative sectional curvature and $C \subset M$ is a compact connected domain with weakly convex boundary then the subsets $C_t = \{x \in C \mid d(x, \partial C) \geq t\}$ are convex in C [18, Chapter 8]. If the second fundamental form of ∂C is

$\geq \frac{1}{r}$ at each point of ∂C, then for all $x \in C$ we have $d(x, \partial C) \leq r$, since the first focal point of ∂C along any inward pointing normal geodesic occurs at distance $\leq r$.

Finally, we recall the statement of the Bishop-Gromov volume comparison inequality. Suppose that M is an n-dimensional Riemannian manifold. Given $p \in M$ and $r_2 \geq r_1 > 0$, suppose that $B(p, r_2)$ has compact closure in M and that the sectional curvatures of $B(p, r_2)$ are bounded below by $K \in \mathbb{R}$. Then

$$\frac{\operatorname{vol}(B(p,r_2))}{\operatorname{vol}(B(p,r_1))} \leq \begin{cases} \frac{\int_0^{r_2} \sin^{n-1}(kr)\, dr}{\int_0^{r_1} \sin^{n-1}(kr)\, dr} & \text{if } K = k^2, \\ \\ \frac{r_2^n}{r_1^n} & \text{if } K = 0, \\ \\ \frac{\int_0^{r_2} \sinh^{n-1}(kr)\, dr}{\int_0^{r_1} \sinh^{n-1}(kr)\, dr} & \text{if } K = -k^2. \end{cases} \tag{G.1}$$

(If $K = k^2$ with $k > 0$ then we restrict to $r_2 \leq \frac{\pi}{k}$.) The same inequality holds if we just assume that $B(p, r_2)$ has Ricci curvature bounded below by $(n-1)\,K$. Equation (G.1) also holds if M is an Alexandrov space and $B(p, r_2)$ has Alexandrov curvature bounded below by K.

Appendix H. Finding balls with controlled curvature

Lemma H.1. *Let X be a Riemannian manifold with $R \geq 0$ and suppose $\overline{B(x, 5r)}$ is a compact subset of X. Then there is a ball $B(y, \bar{r}) \subset B(x, 5r)$, $\bar{r} \leq r$, such that $R(z) \leq 2R(y)$ for all $z \in B(y, \bar{r})$ and $R(y)\bar{r}^2 \geq R(x)r^2$.*

Proof. Define sequences $x_i \in B(x, 5r)$, $r_i > 0$ inductively as follows. Let $x_1 = x$, $r_1 = r$. For $i > 1$, let $x_{i+1} = x_i$, $r_{i+1} = r_i$ if $R(z) \leq 2R(x_i)$ for all $z \in B(x_i, r_i)$; otherwise let $r_{i+1} = \frac{r_i}{\sqrt{2}}$, and let $x_{i+1} \in B(x_i, r_i)$ be a point such that $R(x_{i+1}) > 2R(x_i)$. The sequence of balls $B(x_i, r_i)$ is contained in $B(x, 5r)$, so the sequences x_i, r_i are eventually constant, and we can take $y = x_i$, $\bar{r} = r_i$ for large i. □

There is an evident spacetime version of the lemma.

Appendix I. Statement of the geometrization conjecture

Let M be a connected orientable closed (= compact boundaryless) 3-manifold. One formulation of the geometrization conjecture says that M is the connected sum of closed 3-manifolds $\{M_i\}_{i=1}^n$, each of which admits a codimension-0 compact submanifold-with-boundary G_i so that

- G_i is a graph manifold
- $M_i - G_i$ is hyperbolic, i.e. admits a complete Riemannian metric with constant negative sectional curvature and finite volume
- Each component T of ∂G_i is an incompressible torus in M_i, i.e. with respect to a basepoint $t \in T$, the induced map $\pi_1(T, t) \to \pi_1(M_i, t)$ is injective.

We allow $G_i = \emptyset$ or $G_i = M_i$.

A reference for graph manifolds is [42, Chapter 2.4]. The definition is as follows. One takes a collection $\{P_i\}_{i=1}^N$ of pairs of pants (i.e. closed 2-disks with two balls removed) and a collection of closed 2-disks $\{D_j^2\}_{j=1}^{N'}$. The 3-manifolds $\{S^1 \times P_i\}_{i=1}^N \cup \{S^1 \times D_j^2\}_{j=1}^{N'}$ have toral boundary components. One takes an even number of these tori, matches them in pairs by homeomorphisms, and glues $\{S^1 \times P_i\}_{i=1}^N \cup \{S^1 \times D_j^2\}_{j=1}^{N'}$ by these homeomorphisms. The resulting 3-manifold G is a graph manifold, and all graph manifolds arise in this way. We will assume that the gluing homeomorphisms are such that G is orientable. Clearly the boundary of G, if nonempty, is a disjoint union of tori. It is also clear that the result of gluing two graph manifolds along some collection of boundary tori is a graph manifold. The connected sum of two 3-manifolds is a graph manifold if and only if each factor is a graph manifold [42, Proposition 2.4.3].

The reason to require incompressibility of the tori T in the statement of the geometrization conjecture is to exclude phony decompositions, such as writing S^3 as the union of a solid torus and a hyperbolic knot complement.

A more standard version of the geometrization conjecture uses some facts from 3-manifold theory [59]. First M has a Kneser-Milnor decomposition as a connected sum of uniquely defined prime factors. Each prime factor is $S^1 \times S^2$ or is irreducible, i.e. any embedded S^2 bounds a 3-ball. If M is irreducible then it has a JSJ decomposition, i.e. there is a minimal collection of disjoint incompressible embedded tori $\{T_k\}_{k=1}^K$ in M, unique up to isotopy, with the property that if M' is the metric completion of a component of $M - \bigcup_{k=1}^K T_k$ (with respect to an induced Riemannian metric from M) then

- M' is a Seifert 3-manifold or
- M' is non-Seifert and any embedded incompressible torus in M' can be isotoped into $\partial M'$.

The second version of the geometrization conjecture reduces to the conjecture that in the latter case, the interior of M' is hyperbolic. Thurston proved that this is true when $\partial M' \neq \emptyset$. The reason for the word "geometrization" is explained in [59, 65].

An orientable Seifert 3-manifold is a graph manifold [42, Proposition 2.4.2]. It follows that the second version of the geometrization conjecture implies the first version. One can show directly that any graph manifold is a connected sum of prime graph manifolds, each of which can be split along incompressible tori to obtain a union of Seifert manifolds [42, Proposition 2.4.7], thereby showing the equivalence of the two versions.

References

[1] K. Akutagawa, M. Ishida and C. Lebrun, "Perelman's Invariant, Ricci Flow, and the Yamabe Invariants of Smooth Manifolds", Arch. Math. 88, p. 71-76 (2007)

[2] M. Anderson, "Scalar Curvature and Geometrization Conjectures for 3-Manifolds", in Comparison Geometry, MSRI Publ. 30, Cambridge University Press, Cambridge, p. 49-82 (1997)

[3] M. Anderson, "Scalar Curvature and the Existence of Geometric Structures on 3-Manifolds. I", J. Reine Angew. Math. 553, p. 125-182 (2002)

[4] M. Anderson, "Canonical Metrics on 3-Manifolds and 4-Manifolds", Asian Journal of Math. 10, p. 127-164 (2006)

[5] G. Anderson and B. Chow, "A Pointwise Bound for Solutions of the Linearized Ricci Flow Equation on 3-Manifolds", Calculus of Variations 23, p. 1-12 (2005)
[6] S. Angenent and D. Knopf, "Precise Asymptotics of the Ricci Flow Neckpinch", Comm. Anal. Geom. 15, p. 773-844 (2007)
[7] W. Ballmann, M. Gromov and V. Schroeder, Manifolds of Nonpositive Curvature, Progress in Mathematics 61, Birkhäuser, Boston (1985)
[8] W. Beckner, "Geometric Asymptotics and the Logarithmic Sobolev Inequality", Forum Math. 11, p. 105-137 (1999)
[9] L. Bessières, G. Besson, M. Boileau, S. Maillot and J. Porti, "Weak Collapsing and Geometrisation of Aspherical 3-Manifolds", preprint, http://arxiv.org/abs/0706.2065 (2007)
[10] J.-P. Bourguignon, "Une Stratification de l'Espace des Structures Riemanniennes", Compositio Math. 30, p. 1-41 (1975)
[11] D. Burago, Y. Burago and S. Ivanov, A Course in Metric Geometry, Graduate Studies in Mathematics 33, American Mathematical Society, Providence (2001)
[12] Y. Burago, M. Gromov and G. Perelman, "A. D. Aleksandrov Spaces with Curvatures Bounded Below", Russian Math. Surveys 47, p. 1-58 (1992)
[13] E. Calabi, "An Extension of E. Hopf's Maximum Principle with an Application to Riemannian Geometry", Duke Math. J. 25, p. 45-56 (1957)
[14] H-D. Cao and B. Chow, "Recent Developments on the Ricci Flow", Bull. of the AMS 36, p. 59-74 (1999)
[15] H.-D. Cao and X.-P. Zhu, "A Complete Proof of the Poincaré and Geometrization Conjectures - Application of the Hamilton-Perelman theory of the Ricci Flow", Asian Journal of Math. 10, p. 165-492 (2006)
Erratum to "A Complete Proof of the Poincaré and Geometrization Conjectures - Application of the Hamilton-Perelman theory of the Ricci Flow", Asian Journal of Math. 10, p. 663-664 (2006)
[16] A. Chau, L.-F. Tam and C. Yu, "Pseudolocality for the Ricci Flow and Applications", preprint, http://arxiv.org/abs/math/0701153 (2007)
[17] I. Chavel, Riemannian Geometry - a Modern Introduction, Cambridge Tracts in Mathematics 108, Cambridge University Press, Cambridge (1993)
[18] J. Cheeger and D. Ebin, Comparison Theorems in Riemannian Geometry. North-Holland Publishing Co., Amsterdam-Oxford (1975)
[19] J. Cheeger and D. Gromoll, "On the Structure of Complete Manifolds of Nonnegative Curvature", Ann. of Math. 96, p. 413-443 (1972)
[20] J. Cheeger, M. Gromov and M. Taylor, "Finite Propagation Speed, Kernel Estimates for Functions of the Laplace Operator, and the Geometry of Complete Riemannian Manifolds", J. Diff. Geom. 17, p. 15-53 (1982)
[21] B. Chen and X. Zhu, "Uniqueness of the Ricci Flow on Complete Noncompact Manifolds", J. of Diff. Geom. 74, p. 119-154 (2006)
[22] B. Chow and D. Knopf, The Ricci Flow : An Introduction, AMS Mathematical Surveys and Monographs, Amer. Math. Soc., Providence (2004)
[23] B. Chow, P. Lu and L. Ni, Hamilton's Ricci Flow, Graduate Studies in Mathematics 77, AMS, Providence (2006)
[24] T. Colding and W. Minicozzi, "Estimates for the Extinction Time for the Ricci Flow on Certain 3-Manifolds and a Question of Perelman", J. Amer. Math. Soc. 18, p. 561-569 (2005)
[25] T. Colding and W. Minicozzi, "Width and Finite Extinction Time of Ricci Flow", preprint, http://arxiv.org/abs/0707.0108 (2007)
[26] J. Dodziuk, "Maximum Principle for Parabolic Inequalities and the Heat Flow on Open Manifolds", Indiana Univ. Math. J. 32, p. 703-716 (1983)
[27] J. Eells and J. Sampson, "Harmonic Mappings of Riemannian Manifolds", Amer. J. Math. 86, p. 109-160 (1964)
[28] J.-H. Eschenburg, "Local Convexity and Nonnegative Curvature—Gromov's Proof of the Sphere Theorem", Inv. Math. 84, p. 507-522 (1986)
[29] R. Hamilton, "Four-Manifolds with Positive Curvature Operator", J. Diff. Geom. 24, p. 153-179 (1986)

[30] R. Hamilton, "The Ricci Flow on Surfaces", in Mathematics and General Relativity, Contemp. Math. 71, AMS, Providence, RI, p. 237-262 (1988).
[31] R. Hamilton, "The Harnack Estimate for the Ricci Flow", J. Diff. Geom. 37, p. 225-243 (1993)
[32] R. Hamilton, "A Compactness Property for Solutions of the Ricci Flow" Amer. J. Math. 117, p. 545-572 (1995)
[33] R. Hamilton, "The Formation of Singularities in the Ricci Flow", in Surveys in Differential Geometry, Vol. II, Internat. Press, Cambridge, p. 7-136 (1995)
[34] R. Hamilton, "Nonsingular Solutions of the Ricci Flow on Three-Manifolds", Comm. Anal. Geom. 7, p. 695-729 (1999)
[35] R. Hamilton, "Three-Manifolds with Positive Ricci Curvature", J. Diff. Geom. 17, p. 255-306 (1982)
[36] R. Hamilton, "Four-Manifolds with Positive Isotropic Curvature", Comm. Anal. Geom. 5, p. 1-92 (1997)
[37] H. Ishii, "On the Equivalence of Two Notions of Weak Solutions, Viscosity Solutions and Distribution Solutions", Funkcial. Ekvac. 38, p. 101-120 (1995)
[38] V. Kapovitch, "Perelman's Stability Theorem," in Surveys in Differential Geometry, vol. XI, Metric and Comparison Geometry, eds. J. Cheeger and K. Grove, International Press, Somerville, MA, p. 103-136
[39] B. Kleiner and J. Lott, webpage at http://www.math.lsa.umich.edu/~lott/ricciflow/perelman.html
[40] B. Kleiner and J. Lott, "Locally Collapsed 3-Manifolds", to appear
[41] P. Lu and G. Tian, "Uniqueness of Standard Solutions in the Work of Perelman", preprint, http://www.math.lsa.umich.edu/~lott/ricciflow/perelman.html (2005)
[42] S. Matveev, Algorithmic Topology and Classification of 3-Manifolds, Springer-Verlag, Berlin (2003)
[43] W. Meeks III and S.-T. Yau, "Topology of Three-Dimensional Manifolds and the Embedding Problems in Minimal Surface Theory", Ann. of Math. 112, p. 441-484 (1980)
[44] J. Morgan, "Recent Progress on the Poincaré Conjecture and the Classification of 3-Manifolds", Bull. Amer. Math. Soc. 42, p. 57-78 (2005)
[45] J. Morgan and G. Tian, Ricci Flow and the Poincaré Conjecture, Clay Mathematics Monographs 3, Amer. Math. Soc., Providence (2007)
[46] J. Morgan and G. Tian, "Completion of Perelman's Proof of the Geometrization Conjecture", to appear
[47] G. Mostow, Strong Rigidity of Locally Symmetric Spaces, Annals of Math. Studies No. 78, Princeton University Press, Princeton (1973)
[48] L. Ni, "The Entropy Formula for Linear Heat Equation", J. Geom. Anal. 14, p. 87-100 (2004)
[49] L. Ni, "Ricci Flow and Nonnegativity of Sectional Curvature", Math. Res. Let. 11, p. 883-904 (2004)
[50] L. Ni, "A Note on Perelman's LYH Inequality", Comm. Anal. Geom. 14, p. 883-905 (2006)
[51] G. Perelman, "The Entropy Formula for the Ricci Flow and its Geometric Applications", http://arXiv.org/abs/math.DG/0211159
[52] G. Perelman, "Ricci Flow with Surgery on Three-Manifolds", http://arxiv.org/abs/math.DG/0303109
[53] G. Perelman, "Finite Extinction Time for the Solutions to the Ricci Flow on Certain Three-Manifolds", http://arxiv.org/abs/math.DG/0307245
[54] G. Prasad, "Strong Rigidity of $\mathbb{Q}$-Rank 1 Lattices", Invent. Math. 21, p. 255-286 (1973)
[55] M. Reed and B. Simon, Methods of Modern Mathematical Physics IV, Analysis of Operators, Academic Press, New York (1978)
[56] O. Rothaus, "Logarithmic Sobolev Inequalities and the Spectrum of Schrödinger Operators", J. Funct. Anal. 42, p. 110-120 (1981)
[57] R. Schoen, "Estimates for Stable Minimal Surfaces in Three-Dimensional Manifolds", in Seminar on Minimal Submanifolds, Ann. of Math. Stud. 103, Princeton Univ. Press, Princeton, N.J., p. 111-126 (1983).
[58] R. Schoen and S.-T. Yau, "Complete Three-Dimensional Manifolds with Positive Ricci Curvature and Scalar Curvature", in Seminar on Differential Geometry, Ann. of Math. Stud. 102, Princeton Univ. Press, Princeton, N.J., p. 209-228 (1982)
[59] P. Scott, "The Geometries of 3-Manifolds", Bull. London Math. Soc. 15, p. 401-487 (1983)
[60] V. Sharafutdinov, "The Pogorelov-Klingenberg Theorem for Manifolds that are Homeomorphic to $\mathbb{R}^n$", Siberian Math. J. 18, p. 649-657 (1977)

[61] W.-X. Shi, "Complete Noncompact Three-Manifolds with Nonnegative Ricci Curvature", J. Diff. Geom. 29, p. 353-360 (1989)

[62] W.-X. Shi, "Deforming the Metric on Complete Riemannian Manifolds", J. Diff. Geom. 30, p. 223-301 (1989)

[63] W.-X. Shi, " Ricci Deformation of the Metric on Complete Noncompact Riemannian Manifolds", J. Diff Geom. 30, p. 303-394 (1989)

[64] T. Shioya and T. Yamaguchi, "Volume-Collapsed Three-Manifolds with a Lower Curvature Bound", Math. Ann. 333, p. 131-155 (2005)

[65] W. Thurston, "Three-Dimensional Manifolds, Kleinian Groups and Hyperbolic Geometry", Bull. Amer. Math. Soc. 6, p. 357-381 (1982)

[66] P. Topping, Lectures on the Ricci Flow, LMS Lecture Notes 325, London Mathematical Society and Cambridge University Press, http://www.maths.warwick.ac.uk/ topping/RFnotes.html

[67] R. Ye, "On the l-Function and the Reduced Volume of Perelman I,II", Trans. Amer. Math. Soc. 360, p. 507-531, 533-544 (2008)

[68] R. Ye, "On the Uniqueness of 2-Dimensional κ-Solutions", http://www.math.ucsb.edu/˜yer/ricciflow.html

Department of Mathematics, Yale University, New Haven, CT 06520-8283, USA

E-mail address: `bruce.kleiner@yale.edu`

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109, USA

E-mail address: `lott@umich.edu`